# Bearing fault diagnosis based on a dual-sparse denoising model and convolutional neural network

Xu, Jun; Zhang, Liang; Lu, Guangyin; Liang, Luna; Chen, Jiadui; Sun, Yuanhang; Sun, Jing

**DOI**
10.1088/2631-8695/ae127a

**Publication date**
2025

**Document Version**
Final published version

**Published in**
Engineering Research Express

**Citation (APA)**
Xu, J., Zhang, L., Lu, G., Liang, L., Chen, J., Sun, Y., & Sun, J. (2025). Bearing fault diagnosis based on a dual-sparse denoising model and convolutional neural network. *Engineering Research Express*, *7*(4), Article 045232. https://doi.org/10.1088/2631-8695/ae127a

PAPER

# Bearing fault diagnosis based on a dual-sparse denoising model and convolutional neural network

Engineering Research Express



PAPER

RECEIVED
27 June 2025

REVISED
24 August 2025

ACCEPTED FOR PUBLICATION
13 October 2025

PUBLISHED
27 October 2025

# Bearing fault diagnosis based on a dual-sparse denoising model and convolutional neural network

**Jun Xu**[1], **Liang Zhang**[1,*], **Guangyin Lu**[2], **Luna Liang**[3], **Jiadui Chen**[1], **Yuanhang Sun**[1] and **Jing Sun**[4]

[1] Key Laboratory of Modern Manufacturing Technology of the Ministry of Education, Guizhou University, Guiyang, People's Republic of China
[2] School of Geosciences and Info-Physics, Central South University, Changsha, People's Republic of China
[3] School of Mechanical Engineering, Guizhou University, Guiyang, People's Republic of China
[4] Intelligent Systems Department, Delft University of Technology, Delft, The Netherlands
[*] Author to whom any correspondence should be addressed.

**E-mail:** 1815724800@qq.com, 175050483@qq.com, luguangyin@csu.edu.cn, 634095974@qq.com, jdchen1@gzu.edu.cn, 289956762@qq.com and Jing.Sun@tudelft.nl

**Keywords:** bearing fault diagnosis, signal denoising, deep learning, dictionary learning, wavelet transform

## Abstract

Bearing vibration data are often contaminated with noise, which is detrimental to equipment fault diagnosis and predictive maintenance. Denoising bearing vibration data is crucial. Traditional denoising methods have certain limitations. For instance, when employing wavelet denoising, fixed basis functions may fail to perfectly match all signal structures, potentially compromising denoising accuracy. Similarly, when utilizing data-driven tight frame (DDTF) denoising, the learned basis, due to the lack of noise constraints, may incur a risk of overfitting. To optimize denoising performance in both scenarios, this paper proposes a method that combines wavelet transform and DDTF dictionary learning to extract noise based on a doubly sparse dictionary. The specific approach involves mutually cascading the wavelet transform and DDTF. After applying the wavelet transform to the noisy signal, multi-layer wavelet sparse coefficients are obtained. DDTF processing is then applied to each layer of wavelet sparse coefficients. Subsequent inverse transformation achieves noise suppression. This method integrates the structural constraint capability of wavelet decomposition with the learning capability of DDTF, thereby mitigating their respective limitations to some extent. The denoised data are fed into a residual network model, and training results confirm that the proposed method achieves the best classification performance. Experimental results from both data denoising and deep learning classification demonstrate that the proposed method exhibits superior denoising performance. Although the algorithm structure of this method is more complex compared to other approaches, it is meaningful in scenarios where high-precision denoising is required.

## 1. Introduction

The analysis of bearing signals can provide crucial insights into the bearing's condition. Through an in-depth analysis of these signals, we can detect potential faults or issues in the bearing and take timely measures for maintenance to prevent equipment damage. However, the actual collected bearing data is often affected by various interference noises, making effective denoising of the signals significant for bearing fault detection.

Over the decades, researchers have proposed numerous methods to optimize the denoising of bearing data. Mathematical decomposition is one of the commonly used denoising methods in bearing denoising, such as Empirical Mode Decomposition (EMD) [1, 2]. In this method, high-frequency amplitude modulation signal components generated by localized damage in rolling bearings are decomposed by EMD as Intrinsic Mode Functions (IMFs), then the Hilbert Transform [3] is used to obtain their envelope signals, and the envelope spectrum is calculated to extract the fault characteristic frequencies of the rolling bearing. This method has been applied to analyze the vibration signals of rolling bearings with inner ring damage and outer ring damage

collected on test benches. Due to the adaptive characteristics of the EMD method, it is suitable for the decomposition of non-stationary signals and has been applied to the analysis of fault vibration signals in rolling bearings. However, EMD has the problem of mode mixing, leading to poor denoising performance. To address the mode mixing problem in EMD [4], scholars have improved EMD, such as combining Singular Value Decomposition (SVD) [5] with improved Empirical Mode Decomposition and Variational Mode Decomposition (VMD) [6, 7]. This method first uses SVD to denoise the fault signal to eliminate random interference, then adds high-frequency harmonic signals according to the signal characteristics and decomposes them using EMD, effectively reducing the phenomenon of mode mixing. For VMD, it does not add noise to the IMFs but strictly requires IMFs to be amplitude-modulated and frequency-modulated time series with finite bandwidth based on the Wiener filtering perspective. The number of decomposition layers in VMD can be manually set, and reasonable decomposition layers can be set with the help of the Hilbert Transform envelope of the signal to solve the mode mixing problem. However, VMD is difficult to handle strong human interference, and it is challenging to present reasonable modes [8], reducing the denoising effect.

Apart from the application of mathematical decomposition methods in bearing data processing for denoising, sparse representation is also widely used in this field. For example, Fourier Transform methods [9] are a global transformation method that performs spectral analysis on signals, but lacks the ability to process local detail information when handling signal-noise separation, leading to the loss of a large amount of effective signals. To solve this problem, the Short-Time Fourier Transform denoising method [10] was developed, which has better local description capabilities and can more effectively retain important information in the signal. Although the Short-Time Fourier Transform can achieve local description purposes through time windows, the size of the time window is difficult to adapt [11]. Wavelet Transform [12] can solve this problem well, as its main advantage lies in its good localization properties in both time (space) and frequency domains. In the early stages of its application in the field of bearing signal processing, it gained certain advantages. However, when using Wavelet Transform to process more complex signals, it cannot effectively remove noise, leading to the development of many super wavelet transforms, such as Contourlet Transform [13], which is a multi-scale decomposition method based on super wavelets, particularly suitable for signals with strong directionality and smooth boundaries. Shearlet Transform [14] is a multi-scale transform similar to super wavelet transform but adds a shearing operation to the signal during decomposition, better capturing local features of the signal. Bandlet Transform [15] is a multi-scale transform method used for feature extraction, which decomposes the signal in different directions and processes each sub-band using banding techniques. These methods have all achieved certain effects in denoising processing of signals. Due to the poor adaptability of fixed-base transformations, resulting in poor matching between the mother base and the signal, residual noise is generated, reducing the denoising effect. Single fixed-base transformation is challenging to adaptively represent sparse signals [16]. Therefore, many people have proposed learning-based denoising methods. Learning-based transformations, also known as dictionary learning, are sparse representation methods that construct sparse bases by directly extracting signal features from samples using some optimization algorithm, addressing the drawback that fixed bases cannot change their structure during sparse representation. One typical example is K-Singular Value Decomposition (KSVD), first proposed by Aharon in 2006 [17], which uses SVD [18] for multiple updates to form a sparse base that can adaptively represent sparse signals and achieve better denoising effects. However, this method takes too long for denoising. To solve this problem, many adaptive denoising methods have been proposed, such as DDTF [19, 20], which, compared to KSVD, adds only an orthogonal constraint to the dictionary, allowing DDTF to complete dictionary updates with just one SVD, significantly improving operation speed, and its sparse representation ability is not weaker than KSVD. Of course, deep learning methods also play an indispensable role in bearing fault diagnosis, such as the research by Wang M *et al* on Markov transition field and support vector machine for bearing fault diagnosis [21, 22], and UNet [23] regression for data denoising, among others. Since this study primarily focuses on non-deep learning approaches, these will not be elaborated further here.

To overcome the limitations of single denoising methods, this paper proposes a cascaded denoising framework combining wavelet transform and DDTF dictionary learning. This method first utilizes wavelet transform to perform multiscale sparse decomposition on the noisy signal, obtaining hierarchical wavelet coefficients. Subsequently, at each level, the DDTF algorithm is applied to these sparse wavelet coefficients for further dictionary learning and sparse coding. The processed coefficients are then reconstructed through the inverse wavelet transform to achieve denoising. This dual-sparse dictionary (wavelet + DDTF) design synergistically leverages the structural regularization capability of wavelets and the data-driven learning capability of DDTF, significantly mitigating their respective inherent limitations. The data denoising process using the dual-sparse dictionary is as follows: (1) Determine the number of wavelet decomposition layers K and perform wavelet transform on the input noisy data. (2) Apply a sliding window selection to all obtained high- and low-frequency coefficients, yielding one low-frequency matrix and K high-frequency matrices. (3) Perform DDTF dictionary learning on all high- and low-frequency matrices, where sparse coefficient updating employs the Orthogonal

Matching Pursuit (OMP) algorithm [24], and dictionary updating uses Singular Value Decomposition (SVD). (4) Perform inverse transform and matrix rearrangement on the updated coefficients and dictionaries to extract the noise, followed by an inverse wavelet transform to restore the original data structure. This paper conducts simulated data experiments, real-data experiments, and comparative experiments. The comparative experiments select wavelet denoising, VMD denoising, DDTF denoising, and DDTF-VMD cascaded denoising (applying DDTF processing selectively to the sub-frequencies obtained after VMD) for comparison with the proposed method. Corresponding spectral analysis [25, 26], time-frequency analysis [27, 28], and deep learning-based residual network model evaluations are performed. Through visual result analysis, we verify that the proposed method not only effectively eliminates noise in the data but also outperforms these traditional denoising methods in terms of denoising performance.

## 2. Principle

In the context of bearing vibration signal processing, pertinent signals typically manifest as low-frequency signals or relatively steady-state signals, whereas noise signals commonly exhibit high-frequency characteristics. The principle steps for dual sparse denoising of data signals are as follows:

(1) In the context where a one-dimensional time series $f(t)$ is square-integrable, the definitions of the wavelet transform and its inverse transform are established as follows [29]:

$$Wf(a,\ b) = \int_{-\infty}^{\infty} f(t)\frac{1}{\sqrt{a}}\bar{\psi}\left(\frac{t-b}{a}\right)dt \tag{1}$$

$$f(t) = \frac{1}{C_\psi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\psi\left(\frac{t-b}{a}\right)W_f(a,\ b)\frac{dadb}{a^2} \tag{2}$$

$$C_\psi = \int_{-\infty}^{\infty}\frac{|\hat{\psi}(\omega)|^2}{|\omega|}d\omega < \infty \tag{3}$$

In equations (1)–(3), $\hat{\psi}(\omega) = \int_{-\infty}^{\infty}\psi(t)e^{-i\omega t}dt$ $a$, and $b$ are real numbers. Parameter $a$ represents the scale parameter, determining the frequency of the decomposed signal, while $b$ stands for the location parameter, determining the temporal position of the signal. Function $\psi$ is referred to as the mother wavelet, and $\psi_{a,b}(t)$ as the wavelet, both constructed by scaling and shifting the mother wavelet. Within the theoretical framework of the wavelet transform, the admissibility constant $C_\psi$ plays a dual central role. Physically, it serves as an indispensable normalization factor in the inverse transform formula. Its function is to compensate for the energy distribution characteristics of the mother wavelet itself, ensuring that the signal remains energy-conserved and can be perfectly reconstructed after undergoing the complete process of wavelet decomposition and reconstruction. From the perspective of mathematical rigor, the admissibility condition $C_\psi < \infty$ is itself the fundamental prerequisite for the validity of the inverse transform formula and a guarantee of its convergence. The entire derivation of the inverse transform strictly relies on this integral being finite. Therefore, before employing any function as a wavelet, the primary and essential step is to compute and verify that its $C_\psi$ is finite. This is the cornerstone of wavelet theory, ensuring the mathematical correctness and physical authenticity of subsequent analysis and signal reconstruction. Once the decomposition level K of the wavelet is determined, the data signal is processed using wavelet transformation, yielding K sets of data in different frequency bands.

(2) All obtained high and low-frequency data will be sequentially subjected to a $1 \times 256$ sliding window selection, with each movement step set to $1 \times 1$. This process yields a series of data sets, which are then arranged into data blocks, resulting in a matrix with 256 rows. Similarly, the data obtained from Step ① undergoes identical treatment. Consequently, one low-frequency matrix and K high- frequency matrices are obtained. This procedure constitutes the process of matrix reordering.

(3) For all high and low-frequency matrices, we employ DDTF dictionary learning, with sparse coefficient updates conducted via OMP, and dictionary updates performed. The principle behind dictionary learning is as follows: it is a method capable of adaptively representing sparse texture features of signals. The optimization process within DDTF:

$$\arg\min_{D,A}\ \|\overline{F} - D^T A\|_{\mathrm{F}}^2,\ s.t.\begin{cases}\|a_i\|_0 \leqslant T_0,\ i = 1,\ 2,\ \ldots,P\\ D^T D = \mathrm{I}\end{cases} \tag{4}$$

In equation (4), $\bar{F}$ denotes the overcomplete sample library of signals with a total of p samples. $D$, $A$, and $T_0$

respectively represent the dictionary, sparse coefficients, and upper bound of non-zero elements [30]. $I$ denotes the identity matrix. The sparse encoding of DDTF is formulated as follows:

$$\tilde{a}_i = \arg\min_{a_i} \|\bar{f}_i - D^T a_i\|_2^2,\ s.t. \|a_i\|_0 \leqslant T_0,\ i = 1, 2, ..,P \tag{5}$$

Let $\tilde{A}$ denote the sparse coefficients obtained during iterative computation. In equation (5), $\tilde{f}_i$ and $\tilde{a}_i$ respectively represent the ith column and ith row of $\bar{F}$ and $\bar{A}$.The process of using OMP optimization to optimize equation (5) is as follows [30].

The selection of the sparsity threshold $T_0$ is crucial for balancing the denoising effect and signal fidelity. In this study, $T_0$ is not a fixed value but is adaptively determined based on the noise level. Specifically, we adopt the universal threshold criterion: $T_0 = \sigma\sqrt{2\ln(N)}$, where N = 256 is the length of the sliding window, and $\boldsymbol{\sigma}$ is the standard deviation of the noise. The noise level $\boldsymbol{\sigma}$ is estimated by calculating the median absolute deviation of the highest-frequency detail coefficients from the first-level wavelet decomposition of the noisy signal [31]. An example of determining this parameter through simulation experiments is provided (6), (7), (8):

$$\text{MAD=median}(|\text{cD}_1 - \text{median}(\text{cD}_1)|) = 2.86 \tag{6}$$

$$\hat{\sigma} = \frac{\text{median}(|\text{cD}_1 - \text{median}(\text{cD}_1)|)}{0.6745} = 4.24 \tag{7}$$

$$\text{T}_0 = \sigma\sqrt{2\ln(\text{N})} = 14.12 \tag{8}$$

First of all, in the initial stage, parameter initialization operations need to be carried out. Denote the initial residual as $r_0 = \bar{f}_i$, set the initial indication parameter as $\chi_0$, and set the initial dictionary subset as an empty set, represented by $D_0$. At the same time, set the iteration number $l$ as 0. Next, enter the calculation link. It is necessary to calculate the position parameter $\nu_l = \underset{i \notin \chi_{l-1}}{argmax}\ |\langle r_{l-1}, d_i\rangle|$ of the maximum inner product of the residual r and each column of the dictionary D. Here, $d_i$ represents the i dictionary atom. Then, update the indication parameter $\chi_l = \chi_{l-1} \cup \{\nu_1\}$ according to the obtained position parameter, and expand the dictionary atom $d_{\nu_1}$ at the position corresponding to $\nu_1$ into the dictionary subset $D_l = D_{l-1} \cup d_{\nu_1}$. After that, perform least squares solution on $\bar{f}_i = D_l a_l$ to obtain $\tilde{\alpha}_l = (D_l{}^T D_l)^{-1} D_l{}^T \bar{f}_i$. Here, the superscript T represents matrix transpose. Then, update the residual. The updated residual is represented by $r_l = \bar{f}_i - D_l\tilde{\alpha}_l = \bar{f}_i - D_l(D_l{}^T D_l)^{-1} D_l{}^T \bar{f}_i$. Subsequently, update the iteration number $l = l + 1$ and continue to cycle through the steps. Continue until the condition of $l > T$ is met, that is, when the non-zero element constraint is exceeded, the iteration is stopped. Finally, output R to complete the sparse coding.

For the dictionary update in DDTF, the solution approach differs due to the addition of an orthogonality constraint on the dictionary. The expression for the dictionary update based on DDTF is as follows:

$$\tilde{D} = arg\ \min \|\bar{F} - D^T\tilde{A}\ \|_F^2,\ s.t. D^T D = \text{I} \tag{9}$$

Let $Tr$ represent the trace of the matrix, without influencing the optimization outcome. Therefore, optimizing equation (9) can be equivalently expressed as optimizing the following equation:

$$\tilde{D} = \arg\max_D Tr(D\tilde{A}\bar{F}^T),\ s.t. D^T D = \text{I} \tag{10}$$

In a two-dimensional space, let there be two matrices with identical dimensions, denoted as $\alpha_1$ and $\beta_1$, where the rank of matrix $\beta_1$ is $r_1$. Considering the constraint problem within a compact frame:

$$\tilde{\alpha}_1 = \arg\max_{\alpha_1} Tr((\alpha_1)^T\beta_1),\ s.t. (\alpha_1)^T\alpha_1 = \text{I}_{r_1} \tag{11}$$

If the singular value decomposition (SVD) of $\beta_1$ yields $\beta_1 = U_1\sum_1(V_1)^T$, then $\tilde{\alpha}_1 = U_1(V_1)^T$ holds. Therefore, equation (10) is solvable, with the SVD of $\bar{A}\bar{F}^T = U_\beta\sum_\beta(V_\beta)^T$. By solving the SVD, we can obtain $\tilde{D} = U_\beta(V_\beta)^T$ and $\tilde{D}^T = V_\beta(U_\beta)^T$, thus completing the dictionary update. After implementing the sparse representation based on DDTF, applying the inverse transform on the sparse coefficients and the inverse operation of matrix rearrangement yields one low-frequency and K high-frequency data, resulting in denoised data. Finally, performing the inverse wavelet transform accomplishes denoising [32]. This cascaded approach enables the structural information from the wavelet transform to complement the DDTF, as the wavelet transform provides multi-scale and time-frequency localization capabilities through fixed basis functions, allowing a coarse-grained decomposition of the signal while preserving its primary structural features. In

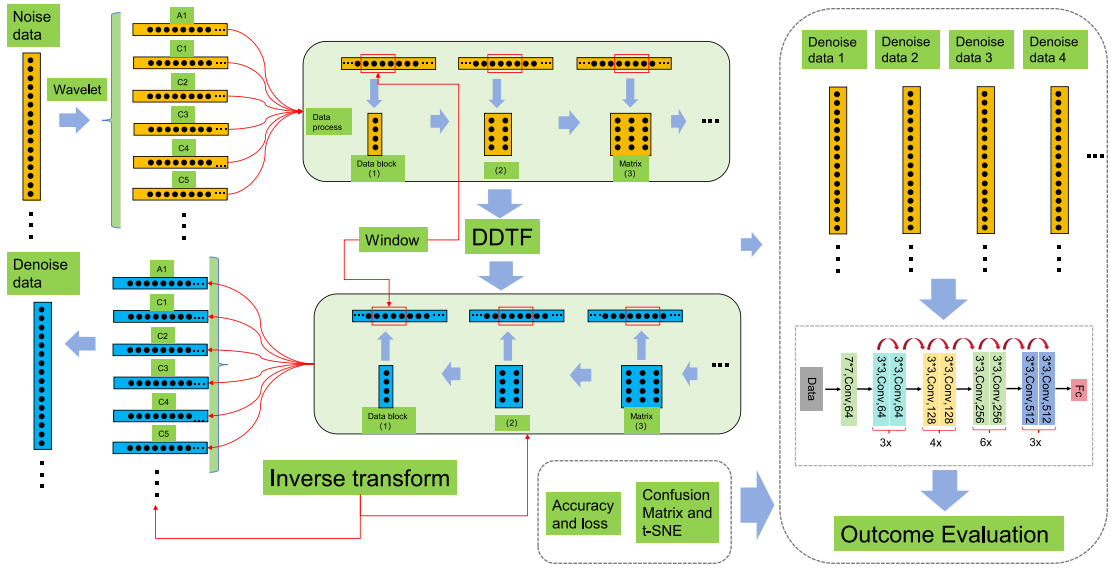


**Figure 1.** The flowchart.

contrast, the DDTF learns data-driven adaptive basis functions to represent fine signal details, yet it lacks global structural constraints and is prone to overfitting to noise. By employing the wavelet transform as a preprocessing step, the main structural components of the signal can be extracted in advance, thereby reducing noise interference during the learning process of the DDTF. Simultaneously, the DDTF performs adaptive sparse reconstruction of the wavelet coefficients, which can further eliminate residual noise while preserving detailed information. This combination effectively leverages the structured representation capacity of the wavelet transform and the adaptive learning ability of the DDTF, thus achieving a balance between global structure and local details in the denoising process.

(4) After denoising the experimental data, the residual network model is employed, and the denoising effect is evaluated through visualization, The flowchart as shown in figure 1.

The present study employs signal-to-noise ratio and Pearson correlation coefficient. analyses to evaluate the denoising efficacy of bearing data.

$$\frac{S}{N} = 10\log_{10}\left(\frac{\|\tilde{Y}\|_F^2}{\|\tilde{Y} - Y\|_F^2}\right) \tag{12}$$

In the equation (12), $\tilde{Y}$ represents raw data.

The Pearson correlation coefficient (PCC) quantifies the linear relationship between two variables by computing their covariance and standard deviations. Its value ranges from −1 to 1, where 1 indicates perfect positive correlation, −1 indicates perfect negative correlation, and 0 indicates no linear relationship. The formula is as follows:

$$r = \frac{\Sigma(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\Sigma(X_i - \bar{X})^2\Sigma(Y_i - \bar{Y})^2}} \tag{13}$$

The meanings of the parameters in equation (13) are as follows: $X_i$ represents the i-th data point of variable $X$. $Yi$ represents the i-th data point of variable $Y$. $\bar{X}$ represents the mean of variable $X$. $\bar{Y}$ represents the mean of variable $Y$.

## 3. Interpretation of experiments

### 3.1. Data interpretation

*3.1.1. HUST dataset*

Since the simulation data was obtained under conditions with high noise levels, a residual impact processing method was applied to the denoising results. The effectiveness of denoising was then evaluated based on this impact. To highlight the comprehensiveness of the experiment, data with lower noise components were selected from Huazhong University of Science and Technology (HUST). This data was used to assess the

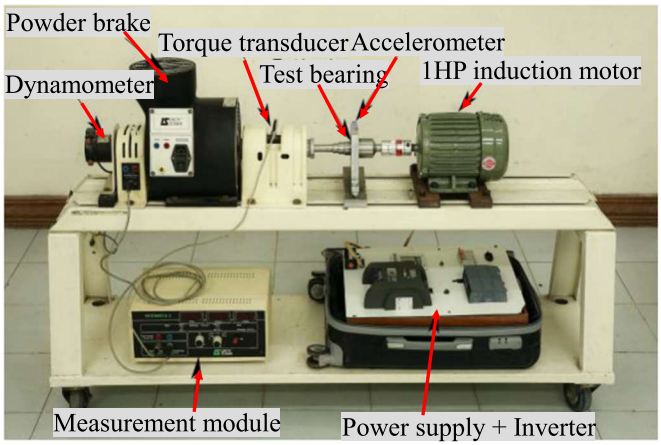


**Figure 2.** HUST experimental platform.

**Table 1.** Structural parameters of bearing.

| Inside diameter | Outside diameter | Ball diameter | Ball number |
|---|---|---|---|
| 20 (mm) | 47 (mm) | 7.6 (mm) | 8 |

**Table 2.** HUST 6-class bearing dataset.

| No. | Data name | Fault type |
|---|---|---|
| 1 | I804.mat | Inner race fault |
| 2 | O804.mat | Outer race fault |
| 3 | B804.mat | Rolling element fault |
| 4 | IB804.mat | Inner race and rolling element faults |
| 5 | IO804.mat | Inner race and outer race faults |
| 6 | OB804.mat | Outer race and rolling element faults |

performance of the proposed method in handling details. Subsequently, high-noise data from Xi'an Jiao tong University (XJTU) was chosen to emphasize the applicability of the results presented in this paper.

To further illustrate the advantages of the proposed denoising method, each algorithm was executed on the publicly available data from HUST [33]. The experiments involved bearings with spherical defects, and the bearing parameters are detailed in table 1, while the test setup is shown in figure 2.

The HUST bearing dataset encompasses vibration signals of bearings under various operational conditions. The health states of the bearings include normal, moderate, and severe faults in the inner race, outer race, rolling element, as well as combined faults. Each state was tested under different operating frequencies, with a sampling frequency of 25.6 kHz.

For the fault classification experiments, the data files were stored in Excel format, with filenames containing fault type and operating frequency information for easy identification and usage. We selected six fault types for classification experiments: inner race, outer race, rolling element, inner and outer race, inner race and rolling element, and outer race and rolling element. The corresponding data are presented in table 2. Additionally, the time-domain waveforms of these six fault types are displayed in figure 3.

### *3.1.2. XJTU dataset*

This study utilizes bearing data from XJTU [34], specifically selecting data pertaining to outer race faults. The experimental bearings used are LDK UER204 rolling bearings, with relevant parameters detailed in table 3.

The rolling bearing test rig used for the experiments in this study is depicted in figure 4. The test rig comprises an alternating current induction motor, a motor speed controller, a support shaft, two support bearings (heavy-duty roller bearings), and a hydraulic loading system. The wear and fracture patterns of the bearing outer race are shown in figures 5(a) and (b).

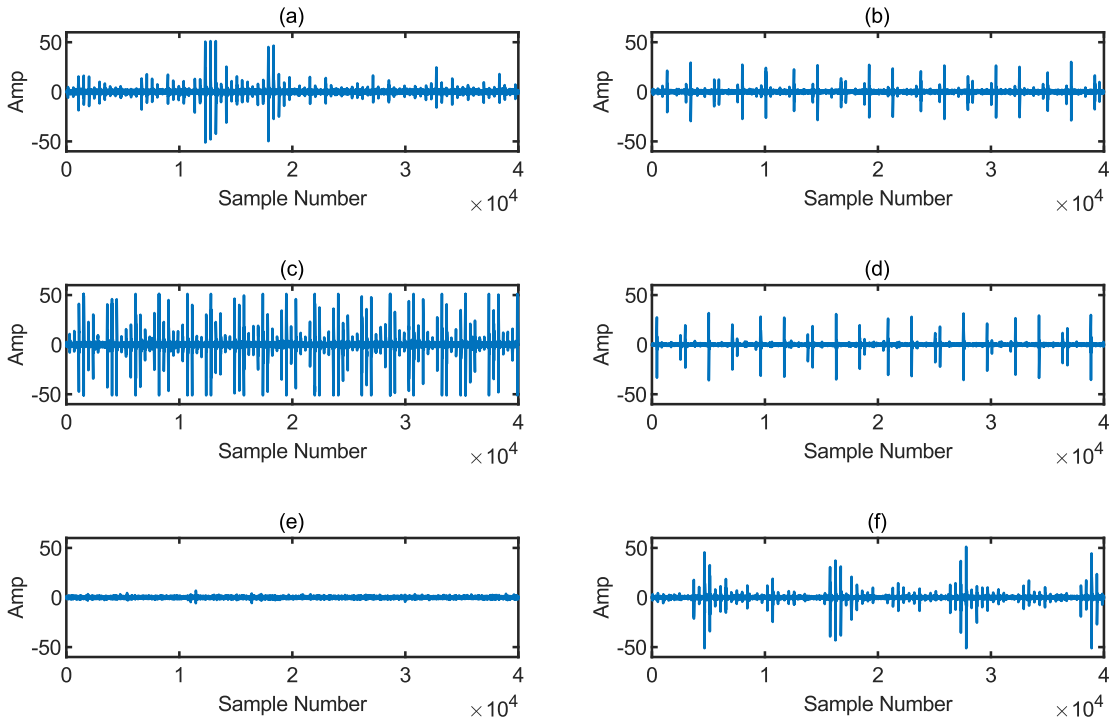


**Figure 3.** The noise signals.

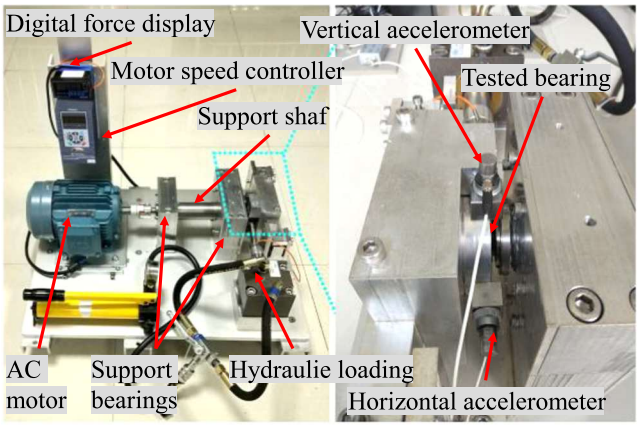


**Figure 4.** The rolling bearing test rig at XJTU.

**Table 3.** Structural parameters of bearing.

| Inside diameter | Outside diameter | Ball diameter | Ball number | Middle diameter | Contact angle |
|---|---|---|---|---|---|
| 29.30 (mm) | 39.80 (mm) | 7.92 (mm) | 8 | 34.55 (mm) | 0° |

### 3.2. Model introduction

As shown in the figure 6, this paper employs a residual network for classification validation experiments. The 34-layer Residual Network (ResNet-34) is a classical deep convolutional neural network whose core architecture addresses the vanishing gradient problem in deep networks through the introduction of residual blocks. The network begins with an initial $7 \times 7$ convolution and a max pooling layer, followed by four main stages in series—comprising 3, 4, 6, and 3 residual blocks, respectively. Each stage consists of multiple stacked residual blocks, with each block containing two $3 \times 3$ convolutional layers and a cross-layer identity shortcut connection that facilitates direct backpropagation of gradients. Finally, classification results are produced via a global average pooling layer and a fully connected layer. In this paper, a 34-layer residual network is employed


**Figure 5.** (a) Wear of the bearing outer ring. (b) The outer ring of the bearing is broken.

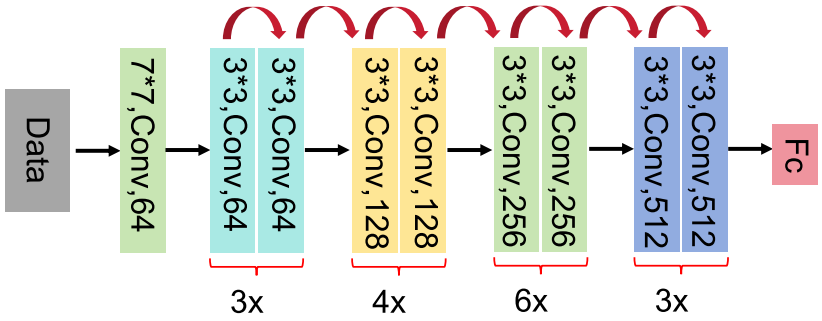


**Figure 6.** ResNet 34.

to classify one-dimensional data, aiming to validate the impact of several data preprocessing methods on experimental results.

### 3.3. Denoising process

The denoising process in this paper is divided into main steps: First, the bearing data signal undergoes K-layer wavelet transform decomposition to obtain signals of different frequencies. Next, these signals are selected via a sliding window and rearranged into matrices, yielding K high-frequency matrices and one low-frequency matrix. Finally, these matrices are processed using DDTF to obtain corresponding dictionaries and sparse coefficients. By applying the inverse transform, the denoised signal segments of different frequencies are reconstructed. Subsequently, an inverse wavelet transform is performed to restore the bearing data to its original form. Through these steps, the noise in the bearing signal is effectively removed. The denoising performance is then evaluated using spectral analysis, time-frequency analysis, Pearson correlation coefficient, signal-to-noise ratio, confusion matrix, and t-SNE visualization.

## 4. Experimentations

In figure 7, in practical data processing, it is often difficult to quantitatively evaluate strong and weak noise based on the signal-to-noise ratio due to the unavailability of clean reference signals. To address this issue, this paper employs a fault impact visualization method based on threshold setting to distinguish and intuitively analyze noise intensity. Specifically, an absolute threshold is applied to the actual data, whereby values below the threshold are set to zero, thereby highlighting the periodic characteristics of fault impacts. Under weak noise conditions, this method clearly reveals the fault frequency, indicating that the influence of weak noise on fault localization is negligible. Denoising analysis and evaluation under weak noise contribute to exploring the potential effect of noise on signal details in low-intensity scenarios. In contrast, under strong noise interference, even with threshold processing, accurately identifying fault features remains challenging, reflecting the higher robustness requirements for such noise. Evaluating denoising performance under strong noise conditions effectively validates the performance of denoising models in extreme environments. For the assessment of denoising, this paper comprehensively adopts methods such as frequency spectrum analysis and time-frequency analysis to systematically evaluate the denoised results.

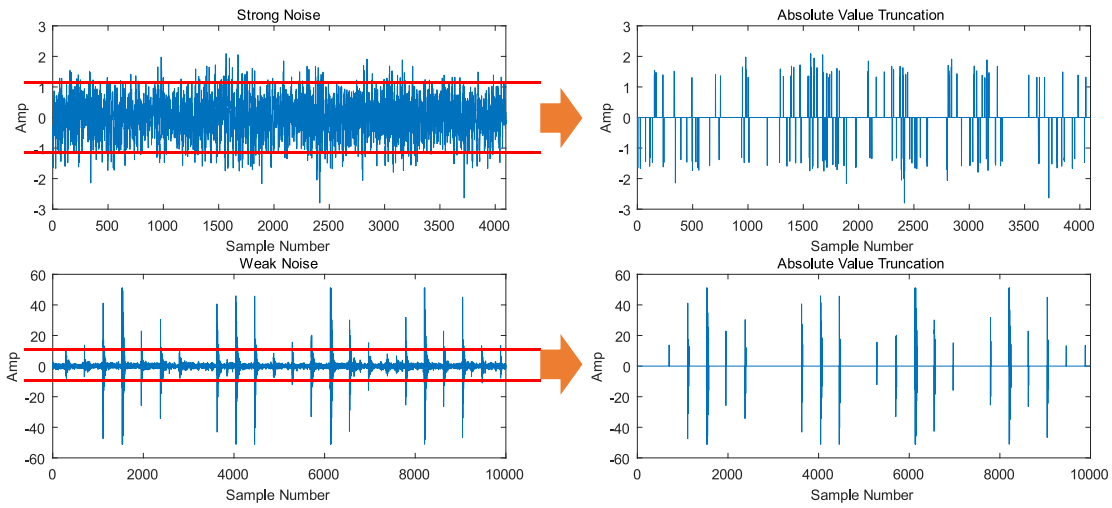


**Figure 7.** Evaluation based on actual data.

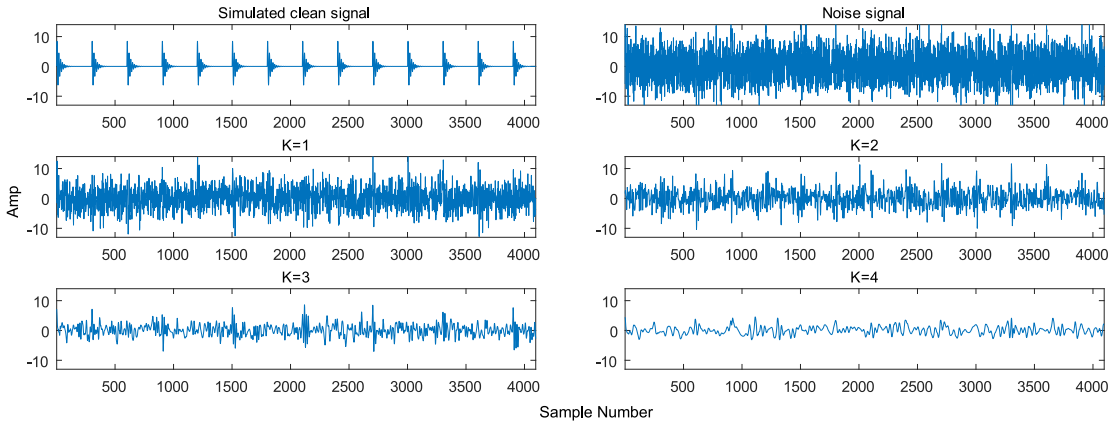


**Figure 8.** The simulated clean signal, −10 dB noise, and the decomposition at K = 1, 2, 3, and 4 levels.

### 4.1. Analog data experimentation

All data preprocessing in this study: DDTF, DDTF-VMD and the proposed method utilize a 1 × 256 time window for sampling noisy data. The specific rationale is as follows: the simulated experimental signal has a length of 4096 (14 cycles), with approximately 290 data points per cycle; the HUST signal has a length of 10000, approximately 22 cycles, with about 454 data points per cycle; the XJTU data has a selected length of 4096, with about 20 cycles and roughly 204 data points per cycle. It is noteworthy that if each window is chosen to be larger than the total data amount of a single cycle, the computational expense would significantly increase. We observe that bearing fault characteristics are similar to those of the clean data in the simulated experiment, with the actual fault-related data points occupying only about one-quarter of each cycle. Therefore, selecting a window size of 256 is reasonable, as it is determined with reference to both the bearing fault characteristics and the fault period. This window size can fully encompass a complete fault signature while considering experimental costs. The experimental environment is: Windows 11, 64-bit operating system, 16 GB RAM, Intel® Core™ 11th Gen i5 processor, NVIDIA GeForce RTX 3050 GPU, MATLAB 2023a, and Python 3.9.0.

The simulated clean bearing signal is shown in figure 8, with the vertical axis representing amplitude (Amp) and the horizontal axis indicating the sample number. It demonstrates that the simulated bearing consistently generates approximately the same vibration amplitude at the fault location. The number of sampling points is set to 4096, covering 14 cycles. The following simulation experiments are conducted with a signal-to-noise ratio of −10 dB after noise addition. The noise-added signal is displayed as the 'Noise Signal' in figure 7. We performed a multi-level wavelet decomposition on this signal and recorded the decomposition results. As observed in figure 6, when K = 1, comparing the simulated clean signal and the noise-added signal, significant noise residue remains at this decomposition level. For K = 3 and K = 4, excessive denoising occurs, resulting in

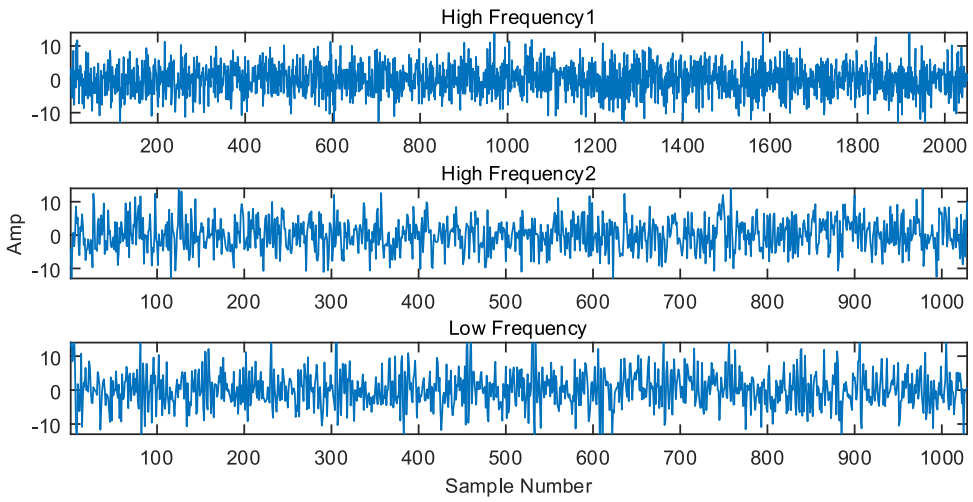


**Figure 9.** Two-layer decomposition of the simulated signal.

the loss of substantial useful information. Overall, K = 2 provides relatively better denoising performance while effectively preserving useful signals. Therefore, we ultimately determined to employ a two-level wavelet decomposition, obtaining two sets of high-frequency coefficients and one set of low-frequency coefficients. It can be observed that the decomposed coefficients contain useful fault impact components, indicating that the subsequent input to DDTF utilizes a reasonable decomposition level. The decomposition results are illustrated in figure 9.

We applied seven denoising methods to the signal respectively and retained the processed impulse components. By comparing figures 10(a) and (b) with figures 10(c)–(i), it can be observed that traditional wavelet denoising and VMD denoising are less effective in noise removal, as they still retain some of the added impulsive noise in non-fault regions. Compared to the clean data, this portion is clearly attributable to noise. DDTF demonstrates improved denoising performance over the former two methods; however, it fails to highlight the impulse features of the original clean signal and still exhibits considerable noise interference in non-fault areas, indicating certain limitations in its denoising effect. Figure 10(f) shows that a significant amount of noise residue remains after processing with this method. The proposed method, shown in figure 10(g), exhibits superior denoising performance compared to the other three methods. The periodicity of the fault signal is clearly visible, with well-preserved impulse effects in each cycle, and the influence of noise is largely eliminated. Figure 10(h) presents a relatively new signal decomposition method, which also shows good denoising results, nearly matching the performance of the proposed method. However, it is noted that the fault location is shifted, which may be related to its inherent algorithm and is not further investigated here. Figure 10(i) uses a self-supervised deep learning denoising method, UNet. From the denoised result, it appears to be the most effective, with the highest signal fidelity. Nevertheless, deep learning methods often suffer from low computational efficiency, making them less suitable for real-time applications. Therefore, the proposed method remains of research interest. A detailed computational efficiency analysis will be provided at the end of this section. Overall, the proposed method offers the best cost-effectiveness.

To further illustrate the superiority of the denoising effect of the method in this paper, we have drawn a signal spectrum diagram for further explanation. As shown in the figure 11(a) represents the clean spectrum diagram of the simulated signal. We can see that its high amplitude is mainly concentrated near 2000Hz and gradually decreases to both sides. It basically decreases to zero at 5000Hz. It can be clearly seen from figure 11(b) the spectrum changes after adding noise. Through figures 11(c)–(g), we can see that the differences in the denoising effect after denoising are mainly reflected in the low-amplitude region and the region within 5000Hz of low frequency. Therefore, we discuss the effects in these two types of regions. After wavelet denoising is used for signal denoising, the denoising effect is obvious when the signal frequency is greater than 5000Hz, but the denoising effect within 5000 Hz and in the low-amplitude part is very poor. The influence of noise in the low-amplitude part is still large, and 5000 Hz retains the possible fault frequency. For VMD denoising, it can be clearly seen that the denoising effect of this method after the frequency is 5000 Hz is very poor, and even an abrupt change in amplitude occurs. The denoising effect near the low amplitude within 5000 Hz is slightly better, but it is also very poor. For DDTF denoising, we can see that the effect in the low-amplitude region within 5000 Hz is better than the first two methods. However, compared with figure 11(a), there is still an amplitude protrusion in the low-amplitude part, and the denoising effect in the low-amplitude region in the

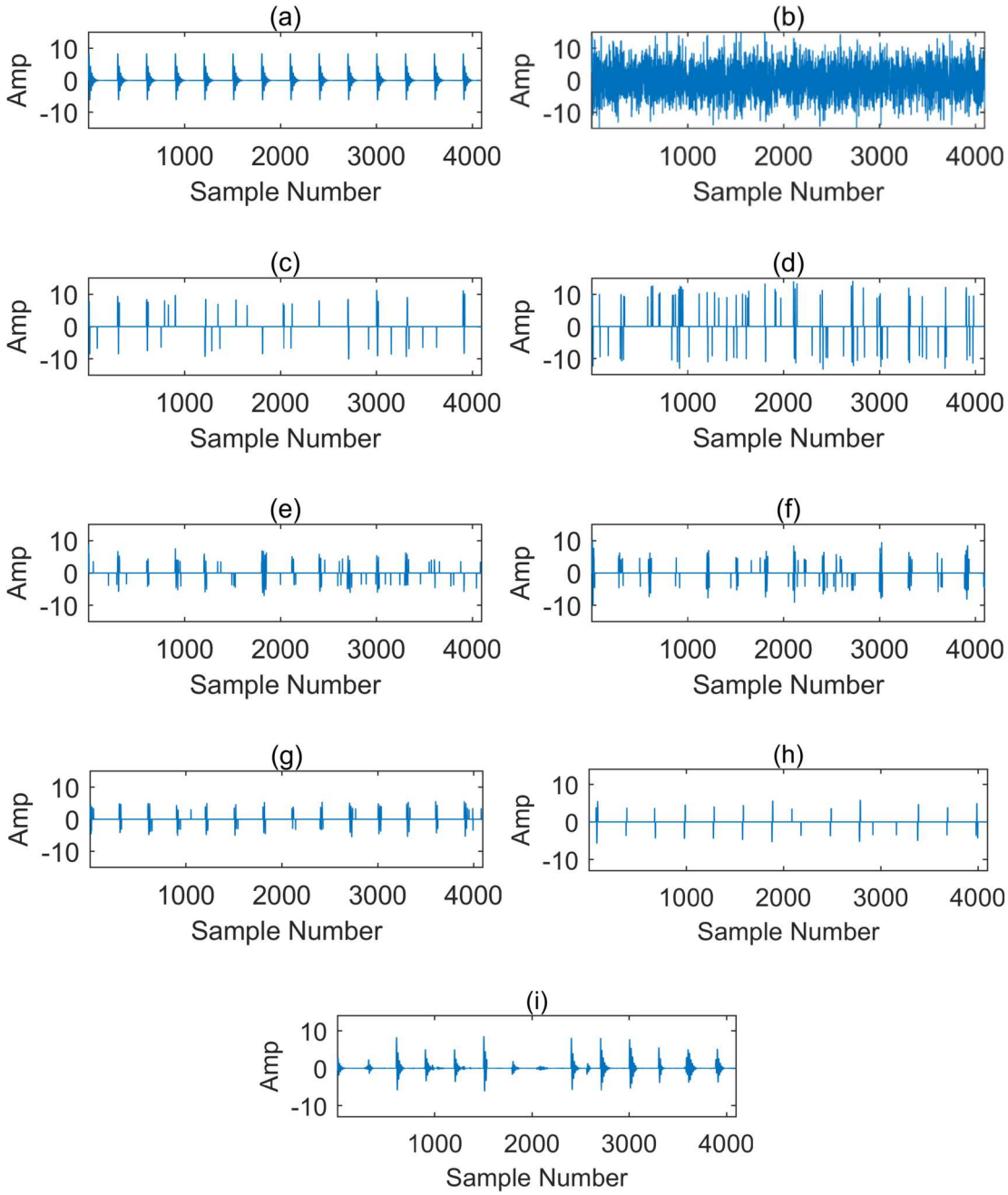


**Figure 10.** Denoising comparison graph. (a) Simulated clean signal. (b) Simulated noisy signal. (c) Wavelet. (d) VMD. (e) DDTF. (f) DDTF-VMD. (g) Our Method. (h) FMD. (i) UNet.

whole domain is not ideal. For the DDTF-VMD method, it can be seen that the spectrum is weakened more seriously. We can clearly see that the denoising effect of the method in this paper in the whole frequency domain is the best. Finally, I performed a local amplification of 0 Hz to 5000 Hz. As shown in figures 11(a1)–(e1), and (g1), it can be seen that the spectrum in this paper is the closest to the simulated clean signal and the detail processing is also relatively the best. Wavelet denoising and VMD denoising are not ideal in denoising in this frequency band. The detail processing of DDTF is not very meticulous. The DDTF-VMD method eliminates a lot of useful signals. Figures 11(h) and (h1) present the spectrum after FMD [35] decomposition and its magnified view, respectively. It can be observed that this method demonstrates effective denoising performance for low-amplitude noise. Figures 11(i) and (i1) display the spectrum denoised using the deep learning method UNet and its enlarged section. The results indicate that this approach achieves excellent denoising, with the restored spectrum closely resembling that of the clean signal. However, despite its effectiveness, the computational efficiency remains an issue that has yet to be fully addressed. In conclusion, the method in this paper is relatively the best.

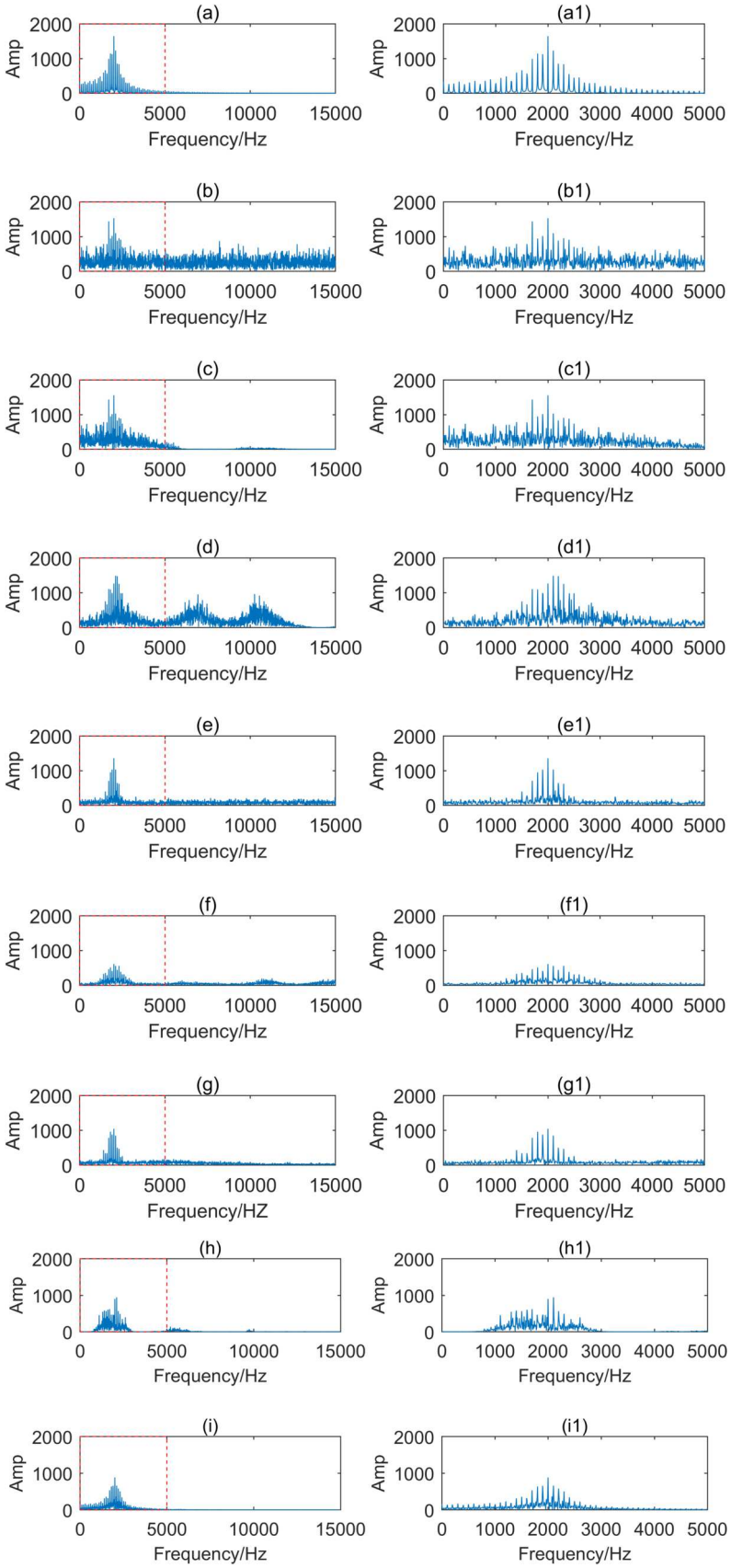


**Figure 11.** Comparison of denoising spectra graph. (a) Spectra of simulated clean signal. (b) Spectra of simulated noisy signal. (c) Wavelet denoising spectrum. (d) VMD denoising spectrum. (e) DDTF denoising spectrum. (f) DDTF-VMD denoising spectrum. (g) Denoising spectrum of the method proposed in this paper. (h) FMD. (i) UNet. (a1) Simulate a locally amplified plot of the clean signal spectrum. (b1) Simulate an amplified local spectrum of a noisy signal. (c1) Wavelet denoising spectrum local amplification diagram. (d1) VMD denoising spectrum local amplification diagram. (e1) DDTF denoising spectrum local amplification diagram. (f1) DDTF-VMD denoising spectrum local amplification diagram. (g1) Our method denoising spectrum local amplification diagram. (h1) FMD. (i1) UNet.

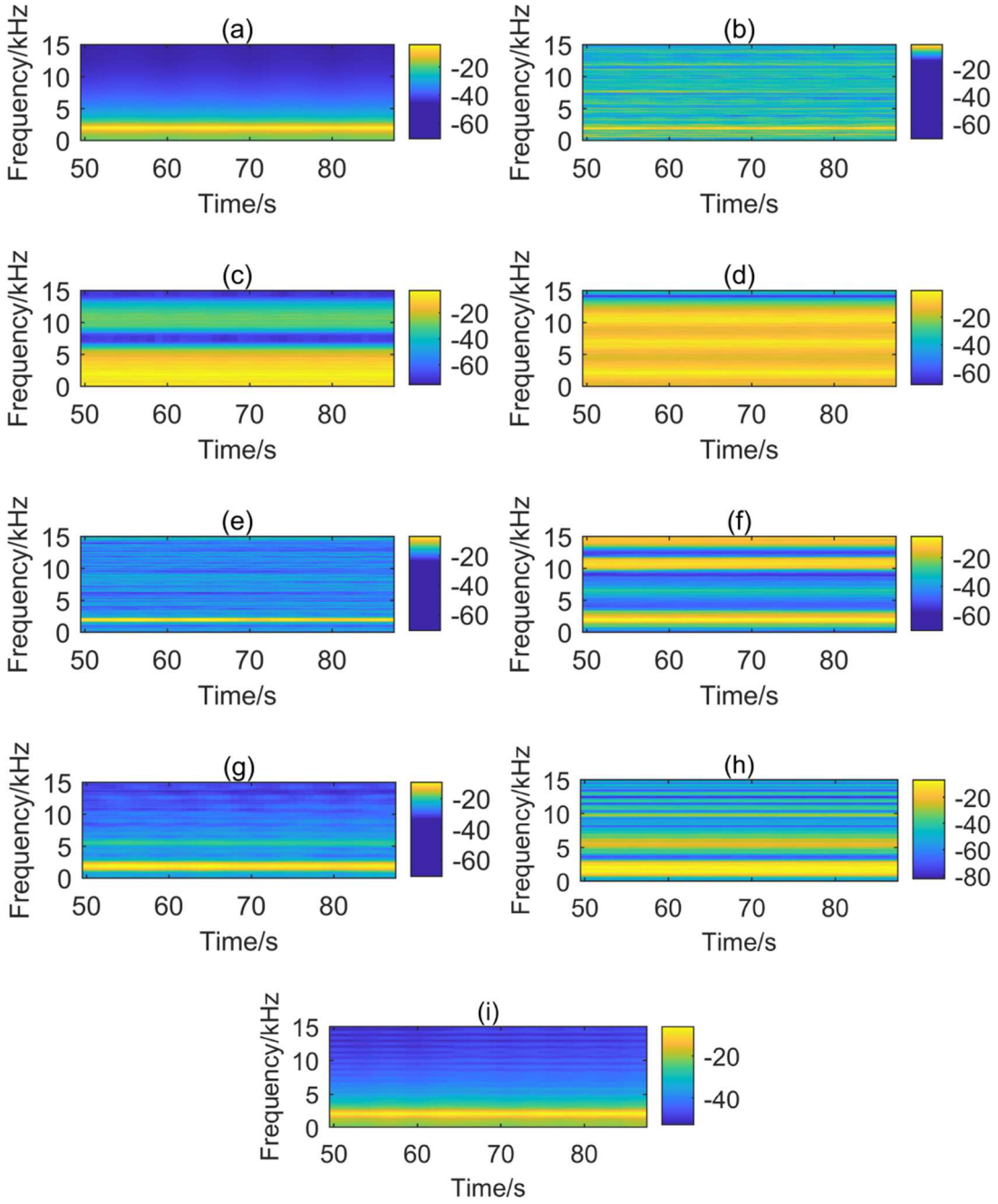


**Figure 12.** Time-frequency diagram. (a) Clean signal. (b) Simulated noisy Signal. (c) Wavelet denoising. (d) VMD denoising. (e) DDTF denoising. (f) DDTF-VMD denoising. (g) Our method. (h) FMD. (i) UNet.

Since the spectrum diagram can only summarize the effects of these methods in the frequency domain, we have also performed time-domain analysis on the data involved in this paper. Time-domain analysis, as a basic and intuitive method, can provide detailed information of the signal in the time dimension. It complements methods such as frequency-domain analysis and provides a comprehensive system understanding and optimization approach. As shown in figure 12(a), by comparing with the spectrum diagram, it can be known that the fault point is probably near 2000 Hz. And in the power time-frequency diagram, the fault location is manifested as relatively high power and is also concentrated near 2000 Hz and decreases to both ends. This is also in line with the signal characteristics. After adding noise, as shown in figure 12(b), the power in the whole domain increases, and the overall power difference is not significant, indicating that the simulated noise is dispersed and has obvious influence. As shown in figure 12(c), the time-frequency diagram obtained after wavelet denoising processing. Combined with the spectrum diagram, it basically shows high power characteristics in the range of 0–5000 Hz. However, the fault frequency of 2000 Hz of the clean signal and the characteristic of decreasing to both ends are not manifested. There is no decreasing effect, indicating that the denoising effect of this method is not good. In figure 12(d), we can clearly see that basically the whole domain has produced high power

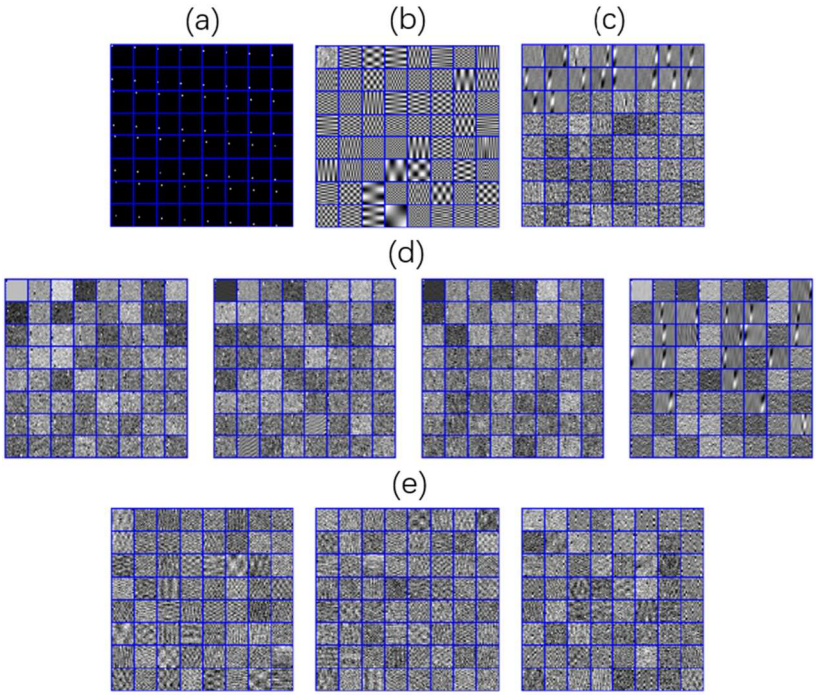


**Figure 13.** Dictionary. (a) The threshold is too high dictionary. (b) The threshold is too low dictionary. (c) DDTF's best dictionary. (d) DDTF-VMD's best dictionary. (e) Text method's best dictionary.

performance. Similar results are also obtained from spectrum analysis. The denoising effect is not good. For the DDTF denoising method, although it is very close to the clean signal in the whole domain, the high-power range at its fault location is relatively small. This is also in line with the spectrum analysis results. However, the overall power is relatively high. There are still some high-power lines in other regions, indicating that the detail processing effect of this method is not ideal. This is similar to the results obtained from spectrum analysis. Figure 12(e) obviously has many obvious high-power points, which further illustrates that VMD and its extended methods are prone to modal aliasing. Looking at the method in this paper again, the power distribution near 2000 Hz is closest to the clean signal, and the effect in the whole domain is also relatively the best. Figure 12(h) displays the time-frequency representation obtained using the FMD decomposition, from which it can be observed that the performance of this method is moderate. Figure 12(i) presents the time-frequency result after denoising with the deep learning-based UNet approach, demonstrating that this method achieves excellent performance and that the reconstructed time-frequency representation closely approximates that of the clean signal. These observations are consistent with the conclusions drawn from the earlier spectral and temporal analyses. Therefore, the method in this paper has the relatively best effect in both time domain and frequency domain.

As shown in figure 13, when we perform denoising with the DDTF method and the method in this paper, we need to set thresholds for the DDTF method, the DDTF-VMD method, and the method in this paper. In figure 13(a), the all-black graph with white dots indicates that when dictionary learning is performed, the threshold is set too large, and some invalid signal features are learned. The striped dictionary graph in figure 13(b) indicates that the threshold is set too small and the signal features cannot be fully learned. Obviously, such threshold selection is unreasonable. The dictionary learned with a reasonable threshold has texture features like those in figurea 13(c)–(e), which also indicates that the thresholds selected in this paper are all relatively optimal.

Furthermore, we recorded the SNR, PCC and computational time for the seven denoising methods applied to the simulated data, as summarized in table 4. As previously mentioned, the initial SNR of the simulation was set to -10 dB. According to equation (13), the corresponding PCC is approximately 0.30. From table 4, it can be observed that the wavelet denoising method achieved a final SNR of −3.68 dB, VMD denoising reached −8.18 dB, DDTF attained −0.91 dB, DDTF-VMD reached −1.41 dB, while the proposed method achieved 4.19 dB. FMD and UNet yielded 1.21 dB and 6.08 dB, respectively. The corresponding PCC values were 0.50, 0.37, 0.45, 0.67, 0.05, and 0.75. However, in terms of computational time, the UNet method required up to 389 s, which is approximately 65 times longer than the proposed method. The proposed method, along with other non-deep learning approaches, required relatively short processing times. This highlights a limitation of deep learning methods, including the time-consuming processes of data processing and training, which make

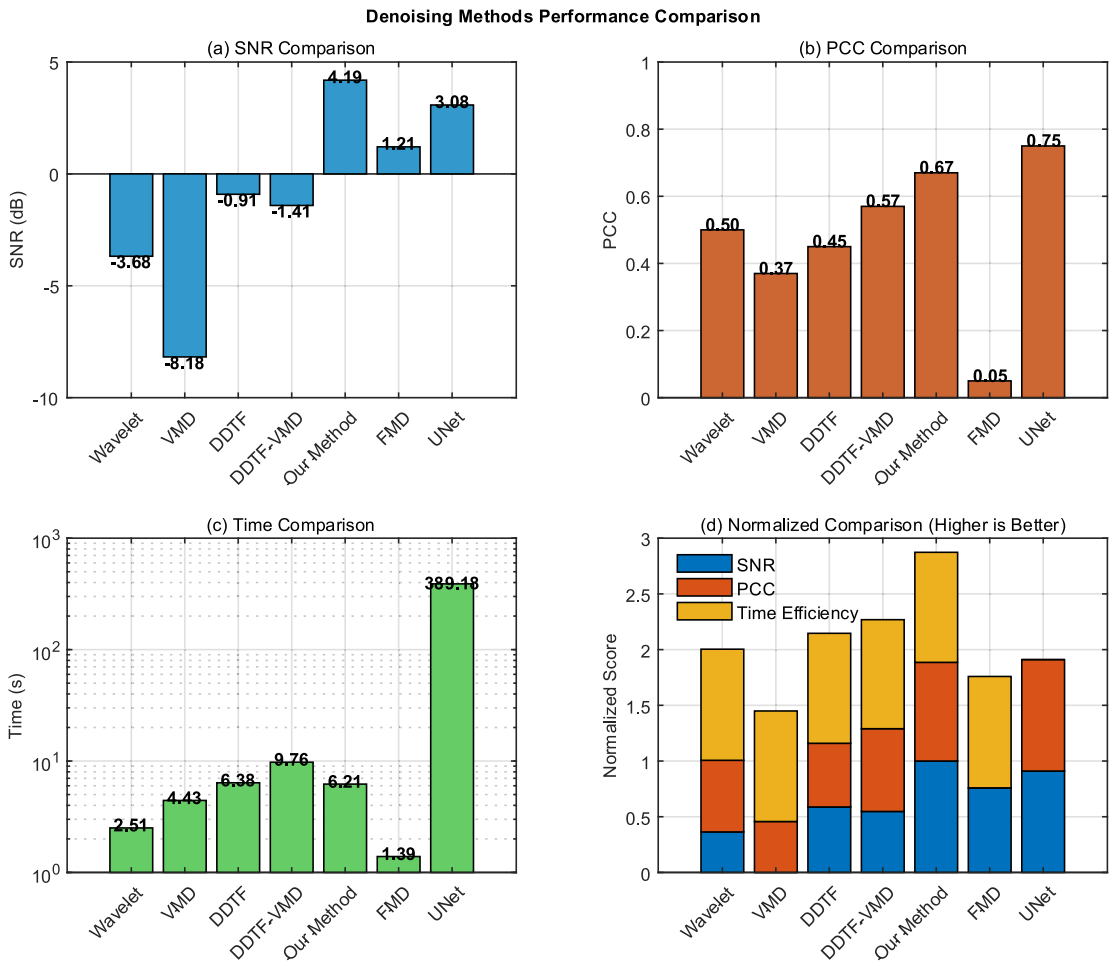


**Figure 14.** Histograms of (a) SNR, (b) PCC, (c) Time, and (d) comprehensive evaluation (higher values indicate better performance).

**Table 4.** Simulated data on SNR, PCC and time.

| Methods | SNR(dB) | PCC | Time(s) |
|---|---|---|---|
| Wavelet | −3.68 | 0.50 | 2.51 |
| VMD | −8.18 | 0.37 | 4.43 |
| DDTF | −0.91 | 0.45 | 6.38 |
| DDTF-VMD | −1.41 | 0.57 | 9.76 |
| Our method | 4.19 | 0.67 | 6.21 |
| FMD | 1.21 | 0.05 | 1.39 |
| UNet | 6.08 | 0.75 | 389.18 |

them less suitable for real-time applications. Therefore, the research on the proposed method remains meaningful, as we explore non-deep learning approaches to improve denoising quality. Integrating SNR, PCC, and computational time, we constructed a bar chart figure 14: (a) compares SNR, (b) compares PCC, (c) compares computational time, and (d) presents an overall cost-effectiveness analysis. Specifically, a higher bar indicates better overall cost-effectiveness. SNR and PCC were normalized to the range [0, 1], where higher values indicate better performance. For the time metric, since lower values are preferable, we used 1 minus the normalized value for evaluation. Thus, the time metric for UNet is nearly zero, indicating excessively long computational time. From this figure, the proposed method demonstrates the best cost-effectiveness performance.

Furthermore, we computed the Pearson correlation coefficients for eight sets of data with initial SNR ranging from −10 to −2 after adding noise. The results are presented in table 5. We also visualized the Pearson correlation coefficients of different methods across various initial SNRs, as illustrated in the line chart in figure 15. It can be observed that the proposed method demonstrates pronounced efficacy when the initial Pearson correlation is low, indicating its robust performance under significant noise influence. When the

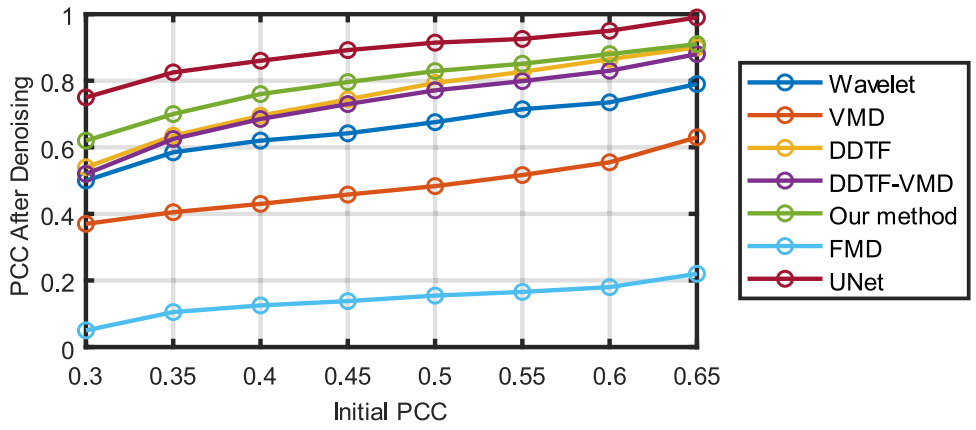


**Figure 15.** The PCC curve.

**Table 5.** Initial PCC and finally PCC.

| InitialPCC | 0.30 | 0.33 | 0.37 | 0.43 | 0.48 | 0.57 | 0.63 | 0.65 |
|---|---|---|---|---|---|---|---|---|
| Wavelet | 0.50 | 0.56 | 0.61 | 0.63 | 0.66 | 0.73 | 0.74 | 0.79 |
| VMD | 0.37 | 0.40 | 0.41 | 0.45 | 0.47 | 0.53 | 0.58 | 0.63 |
| DDTF | 0.45 | 0.48 | 0.67 | 0.72 | 0.78 | 0.84 | 0.89 | 0.90 |
| DDTF-VMD | 0.52 | 0.59 | 0.66 | 0.71 | 0.76 | 0.81 | 0.85 | 0.88 |
| Our Method | 0.67 | 0.69 | 0.73 | 0.79 | 0.84 | 0.87 | 0.91 | 0.93 |
| FMD | 0.05 | 0.09 | 0.12 | 0.13 | 0.15 | 0.17 | 0.19 | 0.22 |
| UNet | 0.75 | 0.81 | 0.84 | 0.88 | 0.91 | 0.93 | 0.97 | 0.99 |

Pearson correlation reaches around 0.65, the final correlation values of the proposed method, the wavelet denoising method, and the DDTF method show little difference, while the UNet approach nearly achieves perfection. This observation is consistent with the earlier experimental conclusions, highlighting the powerful learning capability of deep learning methods. However, considering overall time efficiency, deep learning methods also exhibit temporal limitations. We note that the Pearson correlation of the FMD method remains consistently low. This is because the calculation of the Pearson correlation coefficient requires consideration of positional information in the data, whereas SNR only accounts for power information. As observed earlier in figure 15, the FMD method causes a shift in fault location, though we do not delve into the specific reasons here. The proposed method demonstrates excellent performance both in terms of computational time and effectiveness, indicating that this study also possesses certain practical significance.

### 4.2. Empirical data experimentation

*4.2.1. Weak noise interference, processing real-world data from the HUST*

The original signal effect diagram is shown in figure 16(a). Figure 16(a) contains noise. Due to the extremely large data, we select the first 10,000 data for experiments and analyze the experimental results. The clean fault signal should gradually decrease at the impact position. Figure 16(b) is the signal after wavelet denoising. We can see that the denoising result of the low-amplitude part of its signal is not regular. In figure 16(c), in addition to the fault impact, there is also a large amount of noise at the impact. Some impacts are weakened. In figure 16(d), there is still some noise in its detail part. For example, there is still some low-amplitude noise between impacts that has not been successfully removed, and there is also a certain degree of impact weakening at some periodic positions. And figure 16(e) shows that this method has also weakened useful information to a certain extent, and the detail processing is also not good. The method in this paper is shown in figure 16(f). The fault is obvious, indicating that the denoising effect of the method in this paper is relatively good.

Similarly, we have drawn the spectrum of the original noisy signal. By observing the spectrum of the original noisy signal and comparing and analyzing figures 17(a)–(f), from figure 17(a), we can roughly see the location of the fault frequency. However, there is a great influence of noise in its low-amplitude part. The denoising effects of figures 17(b) and (c) are not obvious. In figures 17(d) and (e), it can be seen that the denoising weakening effect is too strong and useful parts are also weakened. And there are still many noise components in the detail part. the effect of figure 17(f) is relatively the best. Its detail part is relatively the cleanest, and the

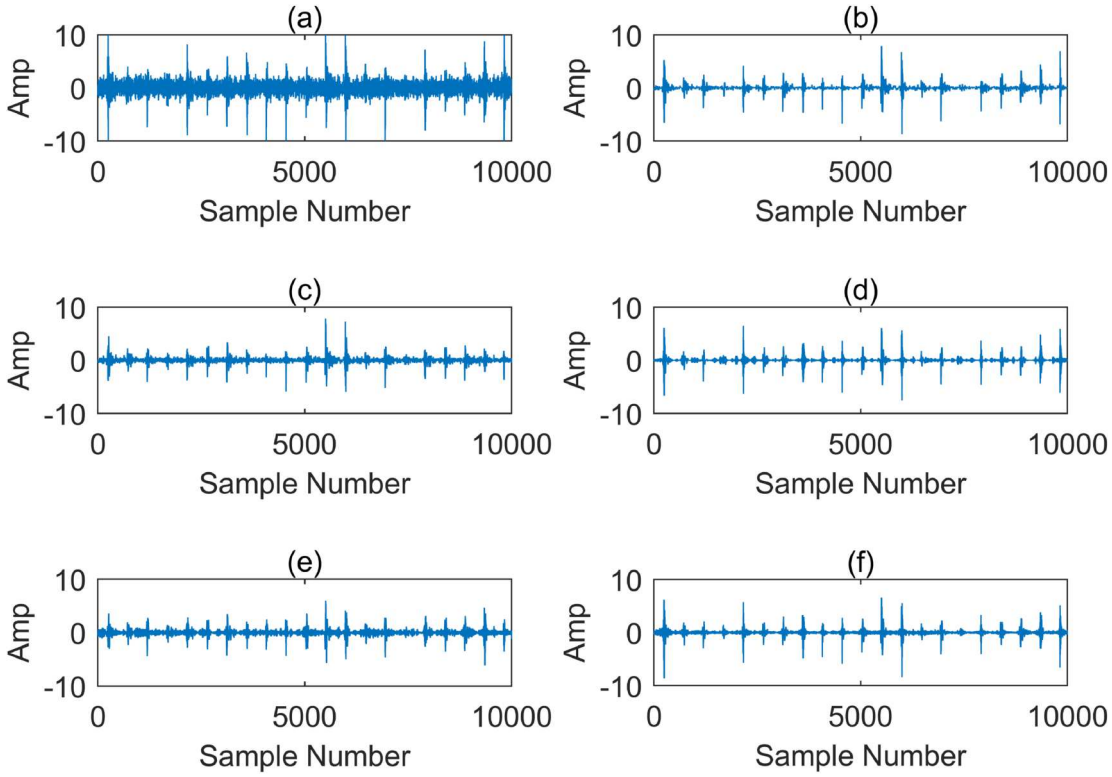


**Figure 16.** Denoising comparison graph. (a) Noisy signal. (b) Wavelet denoising. (c) VMD denoising. (d) DDTF denoising. (e) DDTF-VMD denoising. (f) Our method.

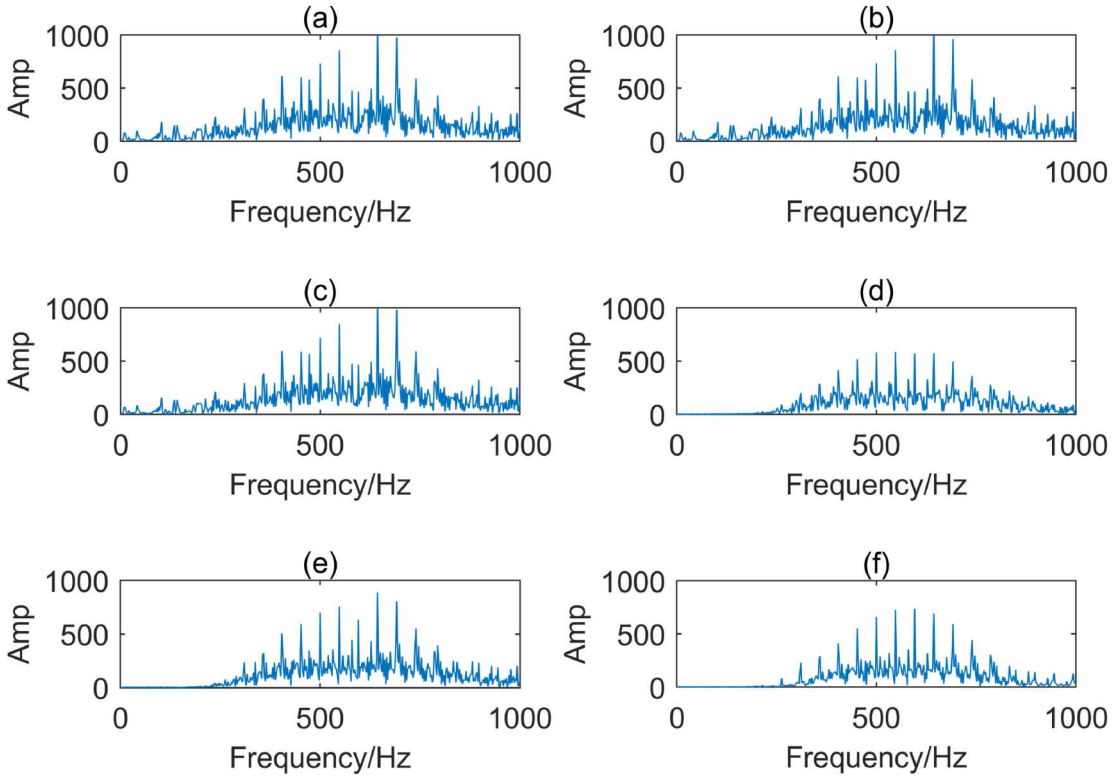


**Figure 17.** Compares the denoising spectra graph. (a) Noisy signal. (b) Wavelet denoising. (c) VMD denoising. (d) DDTF denoising. (e) DDTF-VMD denoising. (f) Our method.

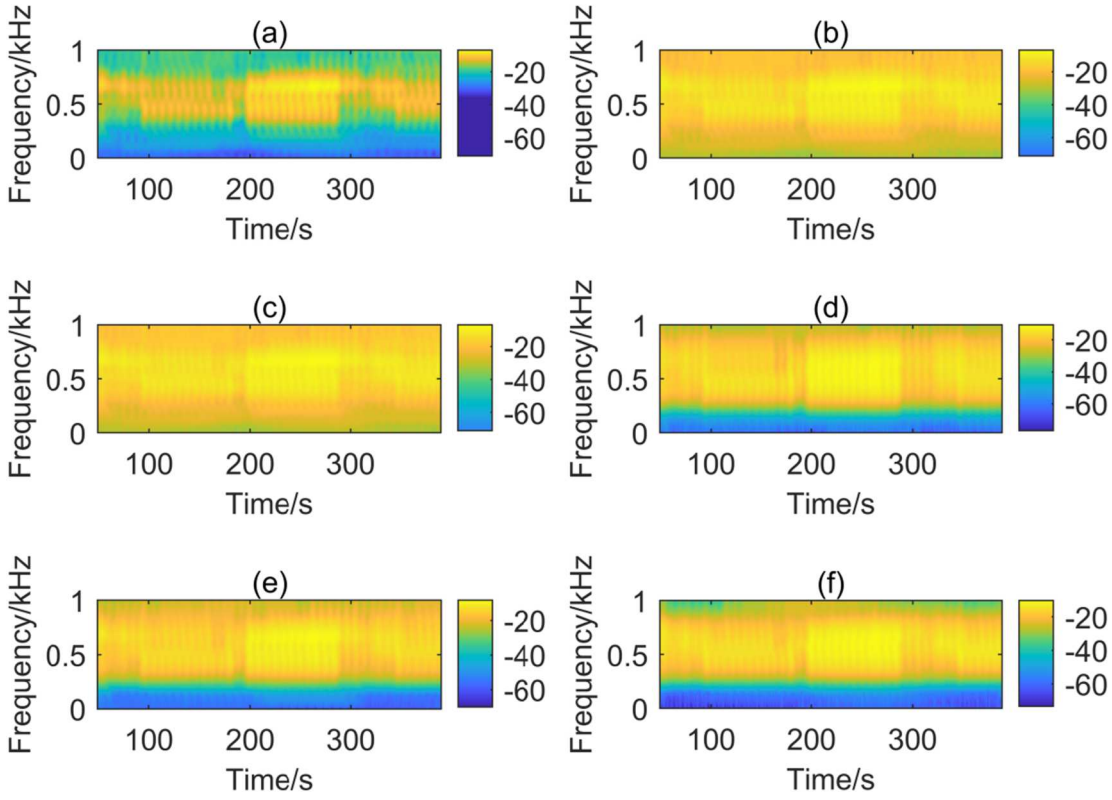


**Figure 18.** Time-frequency diagram. (a) Noisy signal. (b) Wavelet denoising. (c) VMD denoising. (d) DDTF denoising. (e) DDTF-VMD denoising. (f) Our method.

amplitude weakening effect is also the lowest. In conclusion, it also confirms that the effect of the method in this paper is relatively the best.

As shown in figure 18, we also perform time-frequency analysis on these data. We can find that for the wavelet denoising method and the VMD method, the power of the full-domain spectrum is not significantly differentiated, showing very poor denoising performance. While the DDTF method and the DDTF-VMD method have over-denoised, and the effect of the high-frequency part is slightly worse. However, the method in this paper clearly shows the high-power part. This also obtains the same result as the previous signal analysis and spectrum analysis.

*4.2.2. Strong noise interference, processing real-world data from XJTU*

We performed denoising analysis on data selected from XJTU. Due to the high signal-to-noise ratio of the chosen data, the impact of faults was almost entirely obscured by noise. Consequently, we employed impact preservation techniques on the denoised data to validate the efficacy of our proposed denoising method. For signals with strong noise, we focused on analyzing the primary frequency denoising effect. The impact results of the five methods obtained under identical conditions are depicted in figure 19. It is evident that the denoising effect of our method, shown in figure 19(f), demonstrates a more pronounced and cleaner impact pattern compared to figures 19(a)–(e). The other four methods yielded impacts with less regularity and did not provide a clear identification of fault points. Thus, our method exhibits superior performance in this context.

We conducted a comparative analysis of the spectra of four denoising methods relative to the original signal, with the fault frequency corresponding to the fault location provided by XJTU. Besides the fundamental rotational frequency component at 34.38 Hz, the spectrum also exhibited frequency components at 108.6 Hz and its harmonics. This frequency is close to the theoretical fault characteristic frequency of the outer race, which is 107.91 Hz. Figure 20(a) reveals that the original signal contains significant noise in addition to the fault frequency and its harmonics, which is detrimental to fault detection. Figure 20(b) displays the result of wavelet denoising, which fails to effectively extract information in the high-frequency range. Figure 20(c) shows the performance of the VMD denoising method. Although there is some improvement compared to the original signal, the denoising effect remains marginal, with only a minor reduction in noise. Figure 20(d) illustrates a method that performs better than the previous two, although it may have filtered out some useful signals, as the fault frequency location can be preliminarily identified from figure 20(a). Figure 20(e) shows that some frequency doublings of this method are weakened or even completely removed. In contrast, In figure 20(f), the

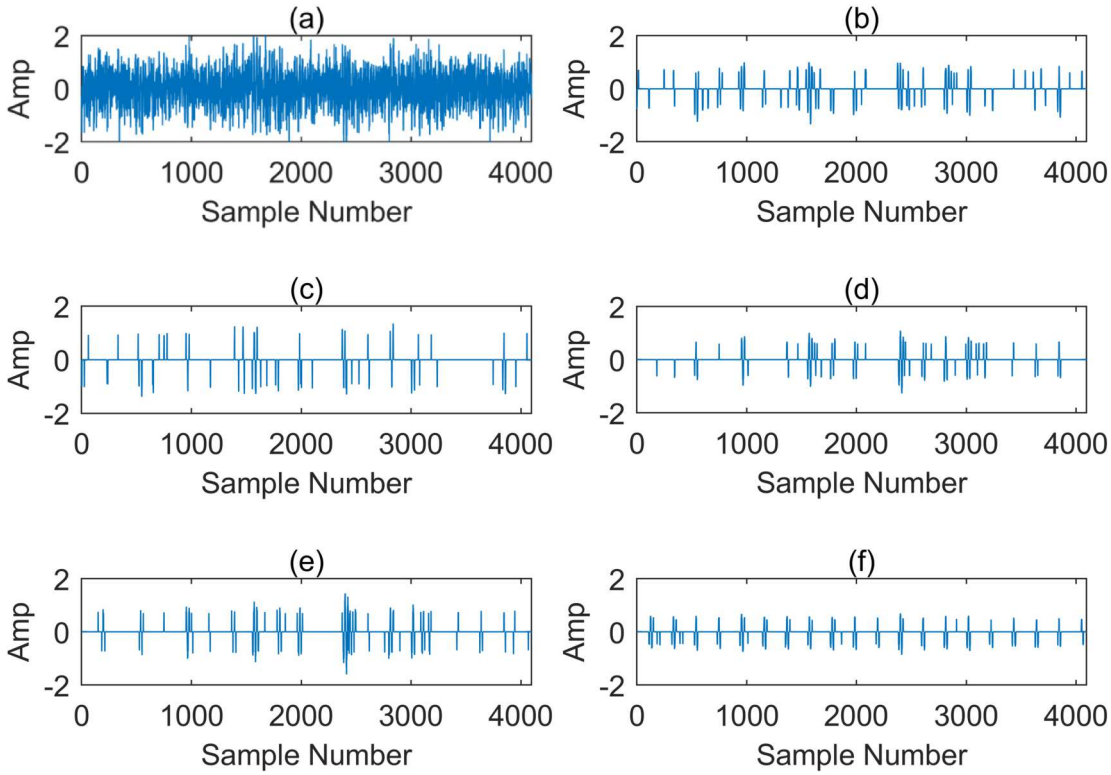


**Figure 19.** Denoising comparison graph. (a) Noisy signal from XJTU. (b) Wavelet denoising. (c) VMD denoising. (d) DDTF denoising. (e) DDTF-VMD denoising. (f) The denoising method proposed in this paper.

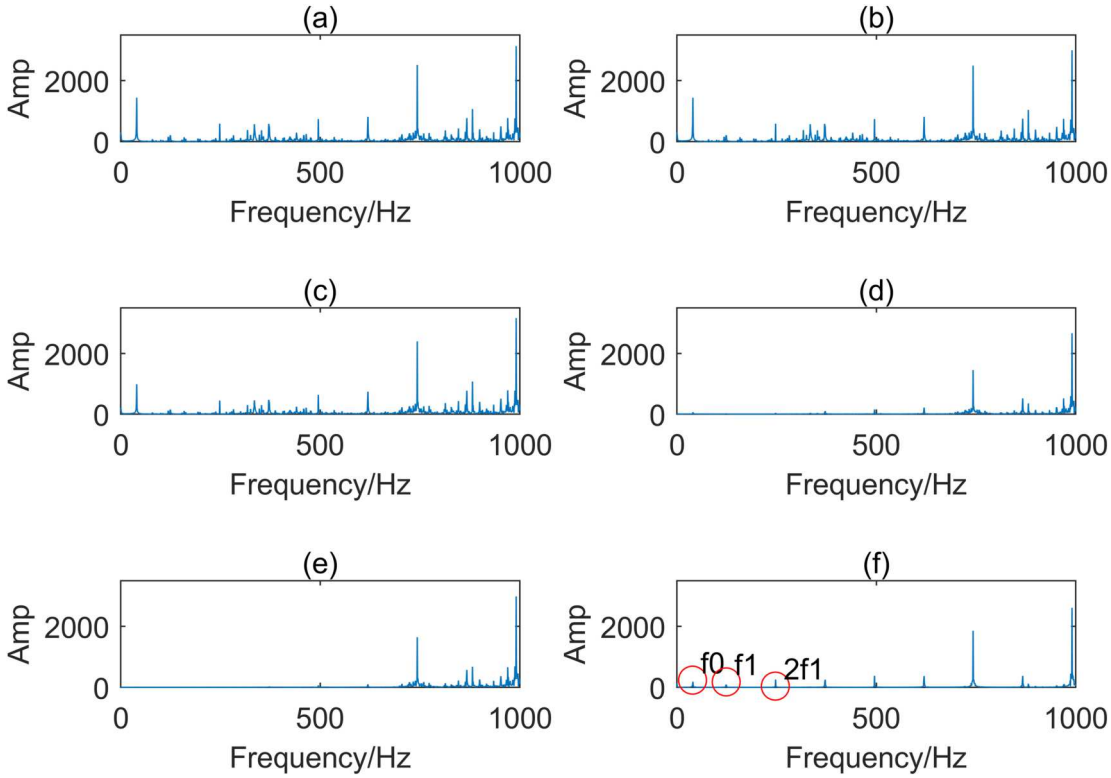


**Figure 20.** Compares the denoising spectra graph. (a) The noisy signal spectrum. (b) The wavelet denoising spectrum. (c) The VMD denoising spectrum. (d) The DDTF denoising spectrum. (e) The DDTF-VMD denoising spectrum. (f) The denoising spectrum using the method proposed in this paper.

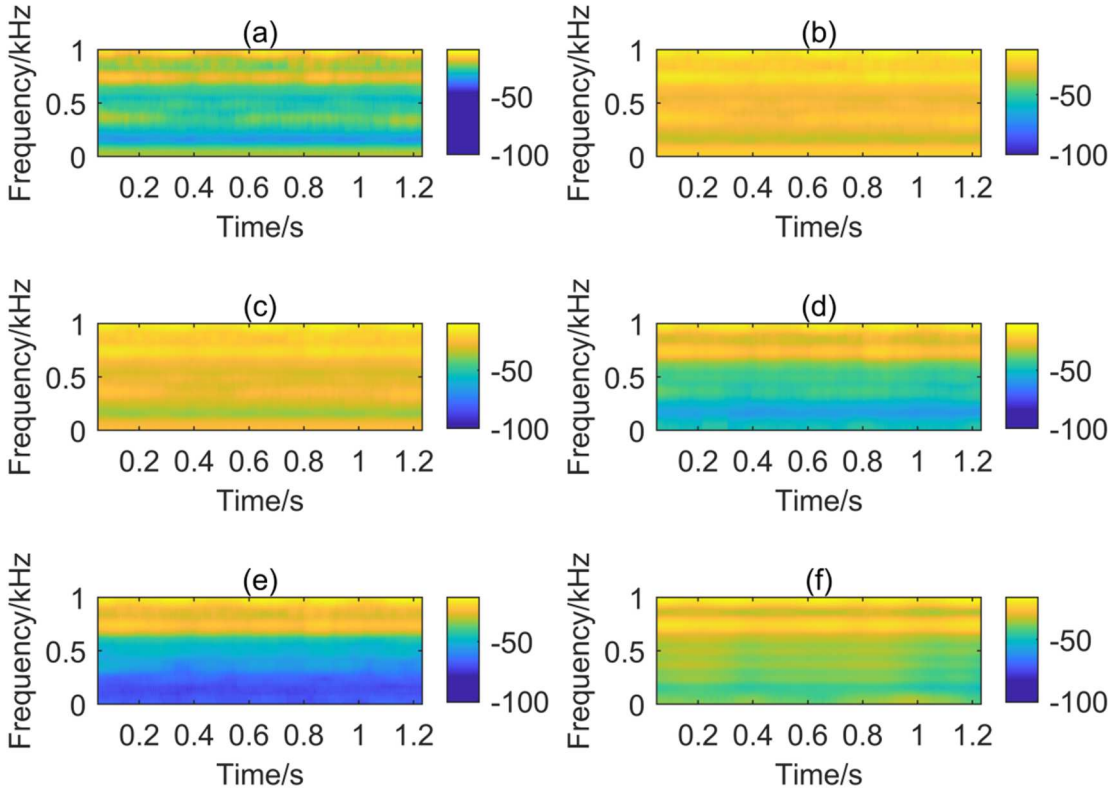


**Figure 21.** Time-frequency diagram. (a) Noisy signal. (b) Wavelet denoising. (c) VMD denoising. (d) DDTF denoising. (e) DDTF-VMD denoising. (f) Our method.

parts circled in red indicate that compared with other methods, the reference rotational frequency and its frequency doublings of this data are obtained more clearly by the method in this paper. In summary, the denoising performance of the proposed method is significantly superior to the other methods, further demonstrating its efficacy in fault signal processing.

Through time-frequency analysis of these data, it can be observed from figure 21(a) that, despite significant noise contamination near these amplitudes, the processing results depicted in figures 21(b) and (c) are not markedly improved. In contrast, the DDTF method shows a more pronounced effect around 1000 Hz, though it still includes considerable noise interference and filters out low-frequency fault components. The DDTF-VMD method reduces a large number of frequency doublings, which is similar to the results of spectrum analysis. The method proposed in this paper demonstrates the best performance, not only making fault frequencies and their harmonics clearly discernible but also effectively reducing the surrounding noise. This further corroborates the reliability of the conclusions.

## 5. Classification experiments based on data denoising

The primary objective of this section is to demonstrate the impact of the denoising method on classification accuracy. Therefore, excessive experiments on model parameter selection and analysis are not conducted. During data segmentation, we process the data using a window size of 256 and a step size of 4. Each dataset is divided into 4,000 samples for the training set and 1,000 samples for the validation set, with no overlap between the training and validation sets to ensure experimental rigor. A 34-layer residual network is adopted as the model architecture, as illustrated in the figure Notably, in the selection of the activation function, as shown in figure 22, bearing fault signals often contain periodic pulse impacts with both positive and negative values. Hence, we employ the ELU activation function instead of the original ReLU to maintain experimental rigor. Through repeated experiments, the learning rate is determined to be 0.000005. Due to the simplicity of the data, the model typically converges after approximately 5 iterations. Thus, to reduce computational costs, setting the epoch to 10 is reasonable.

The table 6 compares the denoising performance of different methods in terms of accuracy and loss, with all values presented as mean $\pm$ standard deviation. The original data (Origin) yields 79.15% $\pm$ 0.3% accuracy and 0.35 $\pm$ 0.01 loss, serving as a baseline. Wavelet denoising performs poorly (61.36% $\pm$ 0.3% accuracy,

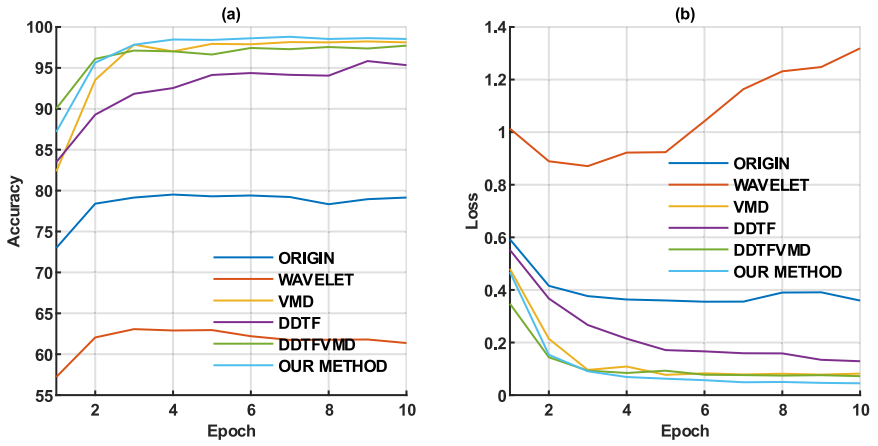


**Figure 22.** Accuracy and loss.

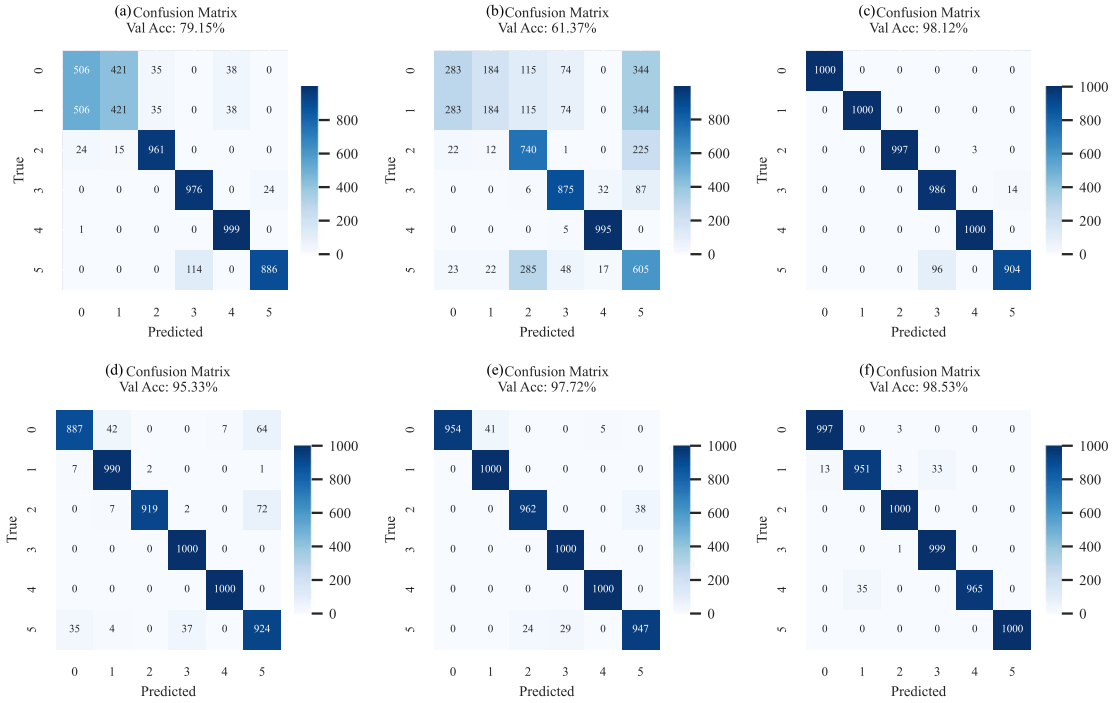


**Figure 23.** Compares the confusion matrix. (a) Without denoising processing. (b) Wavelet. (c) VMD. (d) DDTF. (e) DDTF-VMD. (f) Our method.

**Table 6.** Accuracy and loss rates after data processing by different denoising methods.

| Method | Accuracy | Loss |
|---|---|---|
| Origin | 79.15% ± 0.3% | 0.35 ± 0.01 |
| Wavelet | 61.36% ± 0.3% | 1.31 ± 0.01 |
| VMD | 98.11% ± 0.3% | 0.08 ± 0.01 |
| DDTF | 95.33% ± 0.3% | 0.12 ± 0.01 |
| DDTF-VMD | 97.71% ± 0.3% | 0.07 ± 0.01 |
| Our method | 98.53% ± 0.3% | 0.04 ± 0.01 |

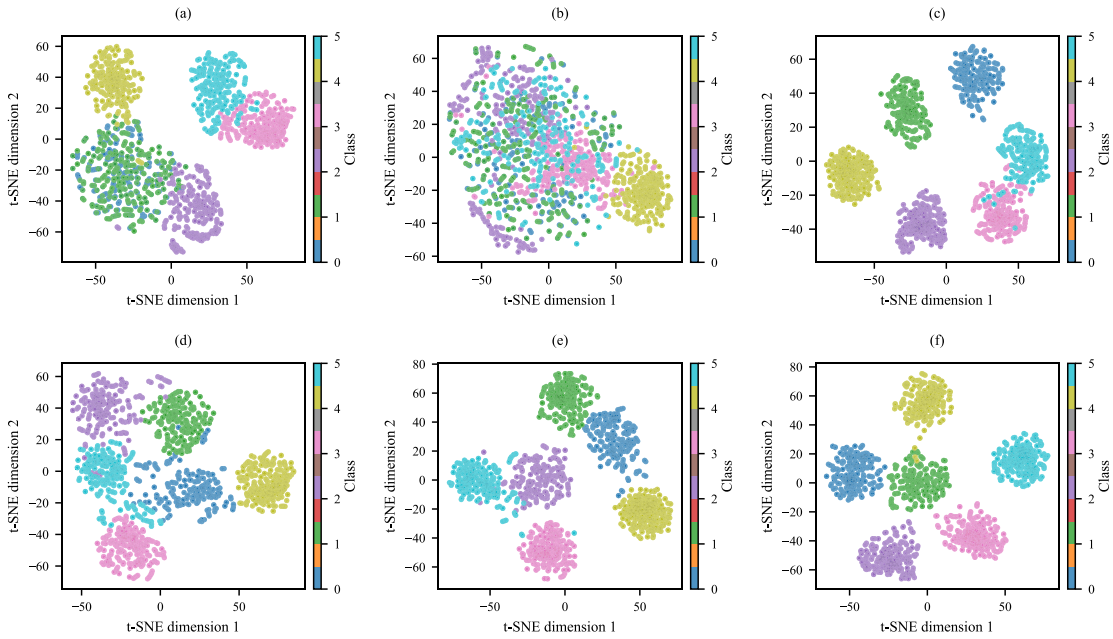


**Figure 24.** Compares the t-SNE. (a) Without denoising processing. (b) Wavelet. (c) VMD. (d) DDTF. (e) DDTF-VMD. (f) Our method.

1.31 ± 0.01 loss), while VMD achieves high accuracy (98.11% ± 0.3%) and low loss (0.08 ± 0.01). DDTF shows competitive results (95.33% ± 0.3%, 0.12 ± 0.01), and the hybrid DDTF-VMD further improves performance (97.71% ± 0.3%, 0.07 ± 0.01). Our method outperforms all others, attaining the highest accuracy (98.53% ± 0.3%) and lowest loss (0.04 ± 0.01), demonstrating its superior denoising capability.

As shown in figure 22(a), which presents the accuracy variation of the validation set when six denoising methods from our experiment were input into the residual network model for six-class classification, our method evidently achieves the highest accuracy. Figure 22(b) demonstrates the loss variation of the six methods during validation, where our method exhibits the lowest loss rate. The trend of these variations clearly validates the effectiveness of our approach.

Figure 23 presents the final confusion matrices of the six methods during validation, demonstrating that our method achieves the best performance. Figure 24 presents a comparative analysis of six t-SNE visualization results. As shown in figure 24(f), distinct boundaries between different data classes are clearly observed, while other methods exhibit varying degrees of overlapping phenomena. This demonstrates that the proposed denoising method achieves the most effective results for these datasets.

## 6. Conclusion

To address the issue of insufficient accuracy in existing methods for denoising bearing vibration signals, the multi-scale dictionary learning method based on wavelet transform proposed in this paper achieves favorable results. We conducted simulation experiments using a self-constructed sample set and introduced real-world datasets for practical validation, with comparative experiments demonstrating the superiority of the proposed method in denoising bearing vibration signals. By comparing wavelet denoising, VMD denoising, DDTF denoising, DDTF-VMD denoising, the proposed method, FMD denoising, and UNet denoising, we conclude that the proposed approach combines the advantages of traditional wavelet denoising and DDTF denoising. It not only reduces the impact of the inflexibility of fixed bases but also mitigates the pseudo-noise artifacts often associated with learned basis denoising methods, thereby preserving useful signals more effectively during noise removal. Although the deep learning method UNet achieves better denoising performance, its high computational time cost makes it unsuitable for real-time applications. The proposed method aims to improve denoising quality without relying on deep learning, offering certain advantages in real-time applicability. Although the computational time of the proposed method is slightly longer than that of wavelet denoising, VMD denoising, DDTF denoising, DDTF-VMD, and FMD denoising, all these methods complete within ten seconds, making the time difference negligible. In contrast, the deep learning method requires 389 seconds, indicating that the proposed approach is more efficient. Despite achieving the best denoising performance and high efficiency, the proposed method involves significantly more parameters to adjust during the tuning

process compared to wavelet denoising and DDTF denoising, and its code complexity is also higher. In the future, the computational redundancy of the proposed algorithm can be further reduced to improve overall quality, which will be the focus of subsequent research.

The experimental code in this paper: https://github.com/YJHHHJ/daima.git.

## Acknowledgments

This work was supported in part by the Deep Earth Probe and Mineral Resources Exploration-National Science and Technology Major Project under Grant No. 2024ZD1002202, in part by Guizhou Provincial Basic Research Program (Natural Science) under Grant No.QianKeHeJiChu -ZK[2024]YiBan013, in part by Guizhou University Basic Research Project under Grant No.GuiDaJiChu[2023]44, and in part by Guizhou University Talent Introduction Research Project under Grant No.GuiDaRenJiHeZi(2023)10.

## Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI:https://github.com/YJHHHJ/daima.git.

## References

[1] Fang K, Zhang H, Qi H and Dai Y 2018 Comparison of EMD and EEMD in rolling bearing fault signal analysis *IEEE Int. Instrum. Meas. Technol. Conf. (I2MTC)* vol. 2018, 1–5 (https://doi.org/10.1109/I2MTC.2018.8409666)
[2] Mundodu Krishna P K and Ramaswamy K 2017 Single channel speech separation based on empirical mode decomposition and Hilbert transform *IET Signal Proc.* **11** 579–86
[3] Kabla A and Mokrani K 2016 Bearing fault diagnosis using Hilbert-Huang transform (HHT) and support vector machine (SVM) *Mech. Ind.* **17** 308
[4] Zhang X, Lv Z and Sun Q 2022 A new method of Aero-engine bearing fault diagnosis based on EMD decomposition *Proc. Prognostics Health Manage. Conf. (PHM-2022 London)* vol., 6–10
[5] Li H, Liu T, Wu X and Li S 2023 Correlated SVD and its application in bearing fault diagnosis *IEEE Trans Neural Netw. Learn. Syst.* **34** 355–65
[6] Guo S *et al* 2025 An optimized VMD and global spatiotemporal model for diagnosing inter-rotor rub-impact faults in dual-rotor turboshaft engine *Eng. Res. Express* **7** 025249
[7] Wang Z *et al* 2019 Application of parameter optimized variational mode decomposition method in fault diagnosis of gearbox *IEEE Access* **7** 44871–82
[8] Jiang F, Zhu Z and Li W 2018 An improved VMD with empirical mode decomposition and Its application in incipient fault detection of rolling bearing *IEEE Access* **6** 44483–93
[9] Sikder N, Bhakta K, Al Nahid A and Islam M M M 2019 Fault diagnosis of motor bearing using ensemble learning algorithm with FFT-based preprocessing *Proc. Int. Conf. Robotics, Electr. Signal Process. Techn. (ICREST)* vol. 2019, 564–9 (https://doi.org/10.1109/ICREST.2019.8644089)
[10] Liu D, Cheng W and Wen W 2020 Rolling bearing fault diagnosis via STFT and improved instantaneous frequency estimation method *Procedia Manuf.* **49** 166–72
[11] Zhang L *et al* 2019 Seismic data denoising via double sparsity dictionary and fast iterative shrinkage-thresholding algorithm *Chinese J. Geophys.* **62** 2671–83
[12] Sardy S, Tseng P and Bruce A 2001 Robust wavelet denoising *IEEE Trans. Signal Process.* **49** 1146–52
[13] Shan H, Ma J and Yang H 2009 Comparisons of wavelets, contourlets and curvelets in seismic denoising *J. Appl. Geophys.* **69** 103–15
[14] Gibert X *et al* 2014 Discrete shearlet transform on GPU with applications in anomaly detection and denoising *EURASIP J. Adv. Signal Process.* **2014** 1–14
[15] McLaughlin M J *et al* 2015 Bandelet denoising in image processing *Proc. Nat. Aerosp. Electron. Conf. (NAECON)* vol. 2015, 35–40 (https://doi.org/10.1109/NAECON.2015.7443035)
[16] Xu J X and Zhang Q 2022 Research and application of an improved wavelet soft threshold function in image denoising *Comput. Eng. Sci.* **44** 1–10 (http://joces.nudt.edu.cn/CN/Y2022/V44/I1/92)
[17] Aharon M, Elad M and Bruckstein A 2006 K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation *IEEE Trans. Signal Process.* **54** 4311–22
[18] Chen Y L and Zhang P L 2012 Bearing fault detection based on SVD and EMD *Appl. Mech. Mater.* **184** 70–4
[19] Zhang L 2019 Research on seismic data reconstruction and denoising methods based on sparse representation *PhD dissertation* Jilin Univ., Changchun, China
[20] Cai J F *et al* 2014 Data-driven tight frame construction and image denoising *Appl. Comput. Harmon. Anal.* **37** 89–105
[21] Wang M *et al* 2022 A new fault diagnosis of rolling bearing based on Markov transition field and CNN *Entropy* **24** 751
[22] Wang M *et al* 2022 Roller bearing fault diagnosis based on integrated fault feature and SVM *Journal of Vibration Engineering & Technologies* **10** 853–62
[23] Li X *et al* 2025 Multimodal U-net: a novel approach for 2D inversion of magnetotelluric data *IEEE Trans. Geosci. Remote Sens.* **63** 1–12
[24] Cai T and Wang L 2011 Orthogonal matching pursuit for sparse signal recovery with noise *IEEE Trans. Inf. Theory* **57** 4680–8
[25] Bian W H and Wang J Y 2024 Multi-channel vibration signal analysis method for rotor-bearing fault *Modeling Simul.* **13** 875
[26] Patil A R, Buchaiah S and Shakya P 2024 Combined VMD-Morlet wavelet filter based signal De-noising approach and its applications in bearing fault diagnosis *Journal of Vibration Engineering & Technologies* **12** 7929–53
[27] Wu X B, Qiu T S and Lü Y S 2004 The development of tiine-frequency distribution based EEG signal analysis and processing *Foreign Med. Biomed. Eng.* **27** 321–6

[28] Liu Y Q, Jing L J and Bian J 2023 Application of time-frequency analysis methods in bearing fault diagnosis *J. Dalian Polytech. Univ.* **42** 1–10
[29] Qiu G X and Chen D H 2003 Application of wavelet transform in signal denoising *J. South China Univ. Technol.* **2003** 13–6 CNKI: SUN:NFYX.0.2003-04-003
[30] Zhang L and Tang J 2024 AMT denoising via double sparse dictionary learning *Journal of Physics: Conf. Series* 2895 (IOP Publishing)
[31] Donoho D L and Johnstone I M 1995 Adapting to unknown smoothness via wavelet shrinkage *J. Am. Stat. Assoc.* **90** 1200–24
[32] Zhang L 2018 Research on seismic data reconstruction and denoising methods based on sparse representation *Master's thesis* Jilin University (https://kns.cnki.net/kcms2/article/abstract?v=Y4WXQ1XfpS44g2w6EtWgXxaWZY4rAeLaQWjRlJB5Be9Nb7QnFA5l3WzkJTOyffQn91II1BmfcPBlkiSqSK2LQ-el-eZNNU19Rxi5mLK3vvwqupFOrIEPpV4iqT_gX1-V0y0XdN2p0d3AMN2LDjNjk553SlHSEHBRwrrDz2Ug_mW0rF_eFHxAp0WtIQVve-no&uniplatform=NZKPT&language=CHS)
[33] Thuan N D and Hong H S 2023 HUST bearing: a practical dataset for ball bearing fault diagnosis *BMC Res. Notes* **16** 138
[34] Lei Y G *et al* 2019 XJTU-SY Rolling Element Bearing Accelerated Life Test Datasets: A Tutorial *J. Mech. Eng.* **55** 1–6
[35] Miao Y *et al* 2022 Feature mode decomposition: new decomposition theory for rotating machinery fault diagnosis *IEEE Trans. Ind. Electron.* **70** 1949–60